Yarmouk University

Community Medicine

Lec. 9 - Measures of disease occurrence

Dr. Ola

Written By : Group F1

If you come by any mistake , please kindly report it to shaghafbatch@gmail.com

Measures of disease occurrence

Lec. 9

MED 410

Dr. Ola Soudah

For a better understanding, please listen to the recordings.

Measures of disease occurrence are needed to estimate incidence and prevalence in certain populations, in order to decide the best study design for these measures, and what factors affect these measures.

Overview

The general rule: Group or part of the population divided by the total number of population.

Proportions

- It is the number of (a) observation with specific characteristics divided by total number of the observations in a given group.

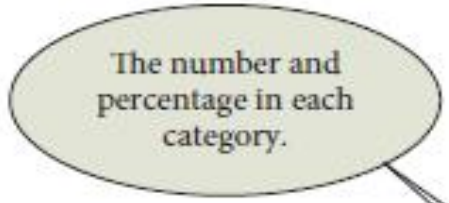
- Example: (a) represent the number of diabetic patients in Jordan , (b) represent the non-diabetics patients in Jordan.

Proportion = a / (a+b)

- Note that:
  - a= diabetics number
  - b= non-diabetics number
  - a+b = n (total population)

Percentage = Proportion x 100

You can multiply the proportion by 100 to get a percentage, or by 10^n



	Cervical pessary group (n = 190)	Expectant management group (n = 190)
Maternal age (years)	30·3 (5·1)	29·6 (5·4)
Body-mass index (kg/m^2)	24·9 (4·6)	24·5 (4·3)
Obstetrical history		
Nulliparous	94 (49%)	96 (51%)
Parous with no previous preterm births	75 (39%)	74 (39%)
Parous with at least one previous preterm birth	21 (11%)	20 (11%)
Cigarette smoking during pregnancy	37 (??%)	38 (??%)
Ethnic origin (self reported)		
White	107 (56%)	110 (58%)
Latin American	58 (31%)	56 (29%)
Other	25 (13%)	24 (13%)
Gestational age at randomisation (weeks)	22·2 (0·9)	22·4 (0·9)
Cervical length at randomisation (mm)	19·0 (4·6)	19·0 (4·9)
Funnelling at randomisation (yes)	81 (43%)	85 (45%)
Sludge at randomisation (yes)	5 (3%)	4 (2%)

Data are number (%) or mean (SD).

Figure 5.1 Baseline characteristics of participants in the cervical pessary and expectant management groups. I have omitted the percentages for cigarette smoking in pregnancy (see Exercise 5.1). Data from Goya *et al.* (2012)

Figure 5.1

Looking at ethnic origin, white people have two groups. Group 1 has cervical pessary and group 2 is the expectant management group, and the total number of white people regardless of the sample is 107+110, which equals 217.

Let's calculate the Column Percent:

The total number of column is (n=190), so $\frac{107}{190} * 100 = 56\%$ which means 56% of cervical pessary group were white people.

Now what is white people percentage who had cervical pessary? Here we calculate Row Percent, so the row total for white people is 217, then $\frac{107}{217} * 100 = 49.3\%$

Rates

The most commonly used measures in descriptive epidemiology to calculate health burden due to certain disease or health outcome.

- Rates are similar to proportions except that the **multiplier** (eg. 1,000 , 10,000 , 100,000) is used, and they are computed **over a specific period of time.**

- The multiplier is called the **base** 10^n Example: 1 per 1000

Rates = Proportion x base or (a / (a + b)) x base

Note that:

a= diabetics number
b= non-diabetics number
a+b = n (total population)

Rates have the same concept as proportions but are time bound.

Example of rates: To figure out the number of Diabetic patients during 2022.

You should know that the more common the disease, the less the base and the rarer the disease, the more the base.

For example:

1- Diabetes is 1 per 5 Persons

2- A certain rare disease is 1 per 100000

Regarding the population in rates: You can restrict the total population at risk. For example: The rates of cervical cancer screening for women who live in Irbid. Here the restriction was by geographical area.

Population at risk

- An important factor in calculating measures of disease frequency is the correct estimate of the numbers of people under study (Denominator).

- Ideally these numbers should only include people who are potentially susceptible to the diseases being studied. For instance, men should not be included when calculating the frequency of cervical cancer.

Figure 2.1. Population at risk in a study of carcinoma of the cervix

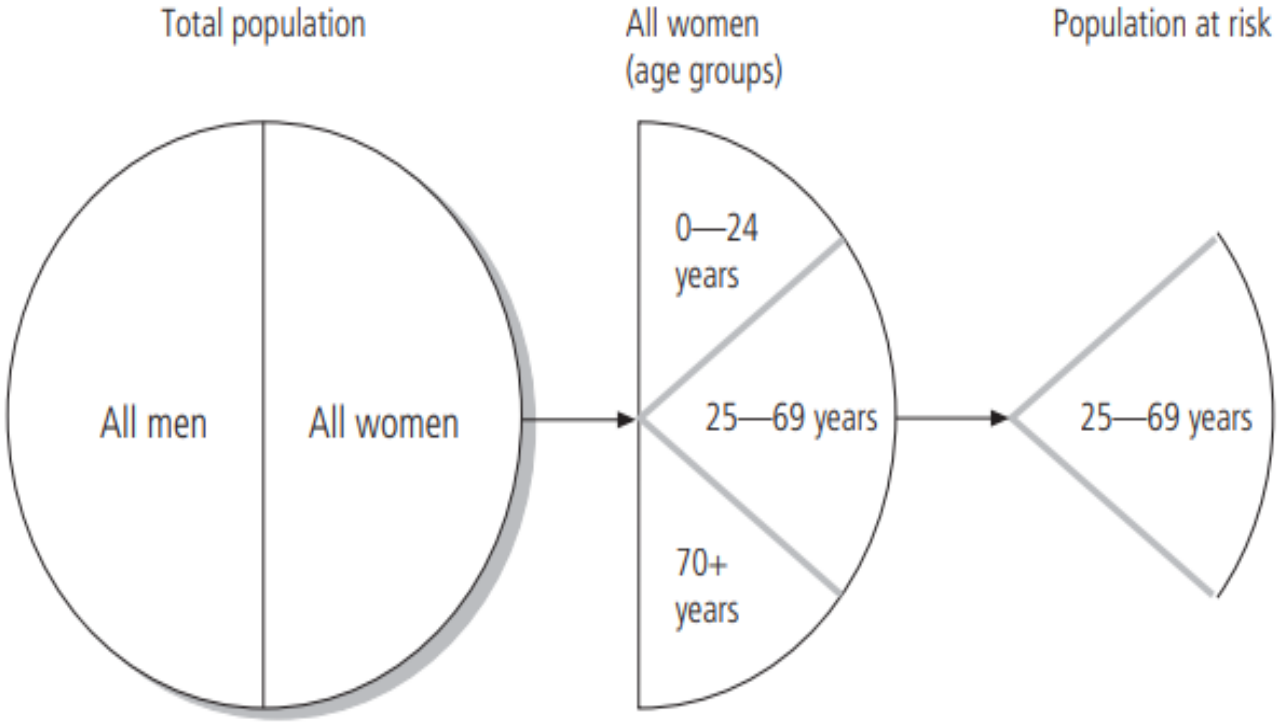


Figure 2.1

The total population of Jordan is 10 million, and the current female population is around 5 million, so to calculate the frequency of cervical cancer in Jordan, we should exclude men, because its only restricted to women. Then we should check the age group at risk. Cervical cancer risk starts at age 26-69. These are the people at risk or the population at risk (Denominator).

Example: Calculate the rate of women who had cervical cancer screening?

$$\frac{Number\ of\ women\ who\ did\ screening}{Total\ population\ at\ risk}$$

Measures of disease Morbidity

Prevalence is a rate because a time factor is included.

Prevalence	Measures **existing** cases of disease and is expressed as a proportion
Incidence	Measures **new** cases of disease and is expressed in person-time units

The incidence is restricted to the newly diagnosed (new cases).

Prevalence

- **Prevalence** is defined as the number of affected persons present in the population at a specific time divided by the number of persons in the population at that time.

Prevalence *(P)* of a disease is calculated as follows:

$$P = \frac{\text{Number of people with the disease or condition at a specified time}}{\text{Number of people in the population at risk at the specified time}} (\times 10^{n})$$

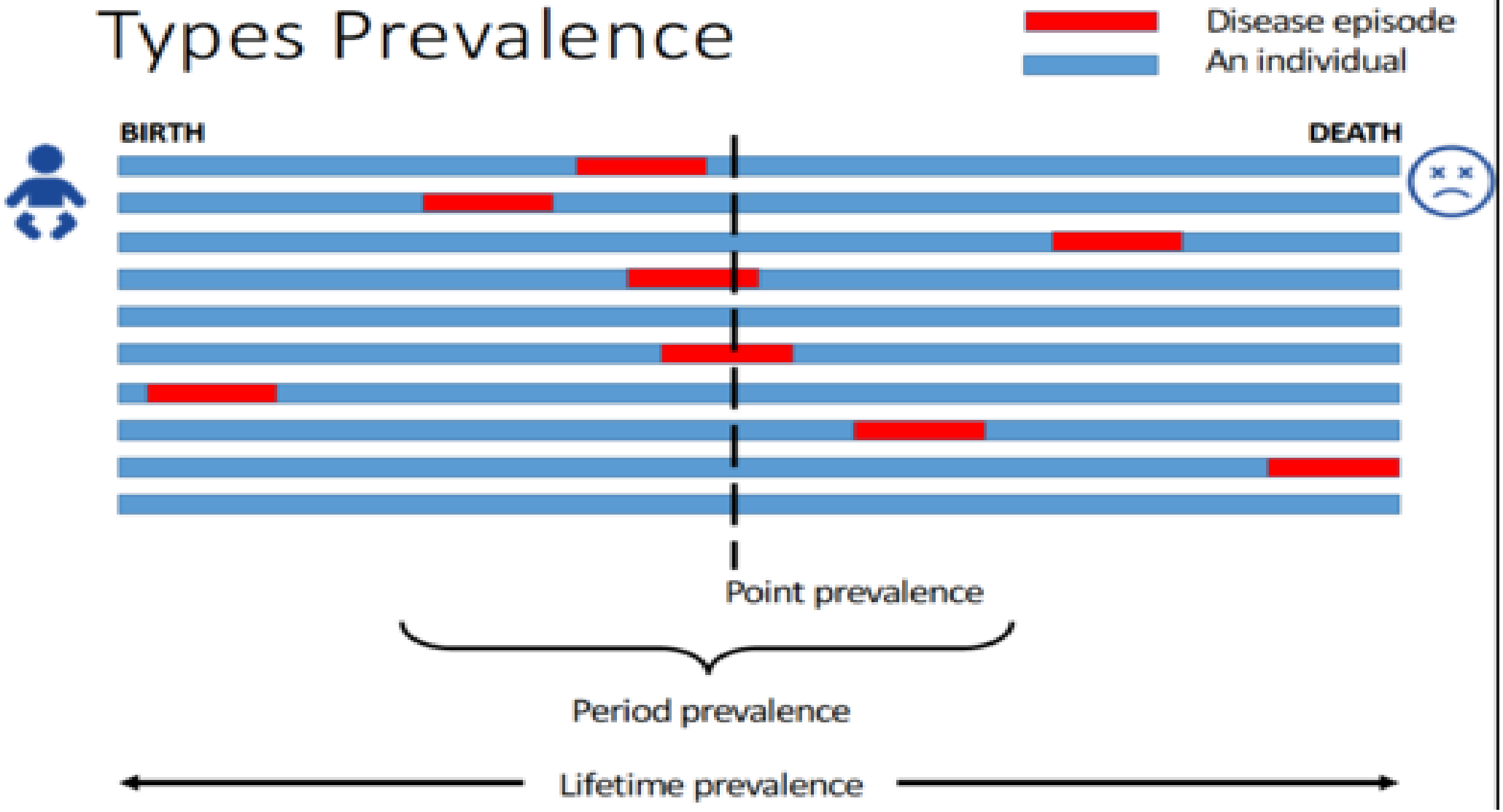
Types of Prevalence

1-At a very specific time.
2-Total population (denominator) is determined by the number people at that time

Point prevalence: number of people with the disease at point of time.

Period prevalence: Number of people with a disease over period of time, the denominator is the average of the population size during that period of time (accounting for dynamic population).

Types Prevalence



For the Figure above:

The prevalence at a specific point of time (Point Prevalence): 2/10

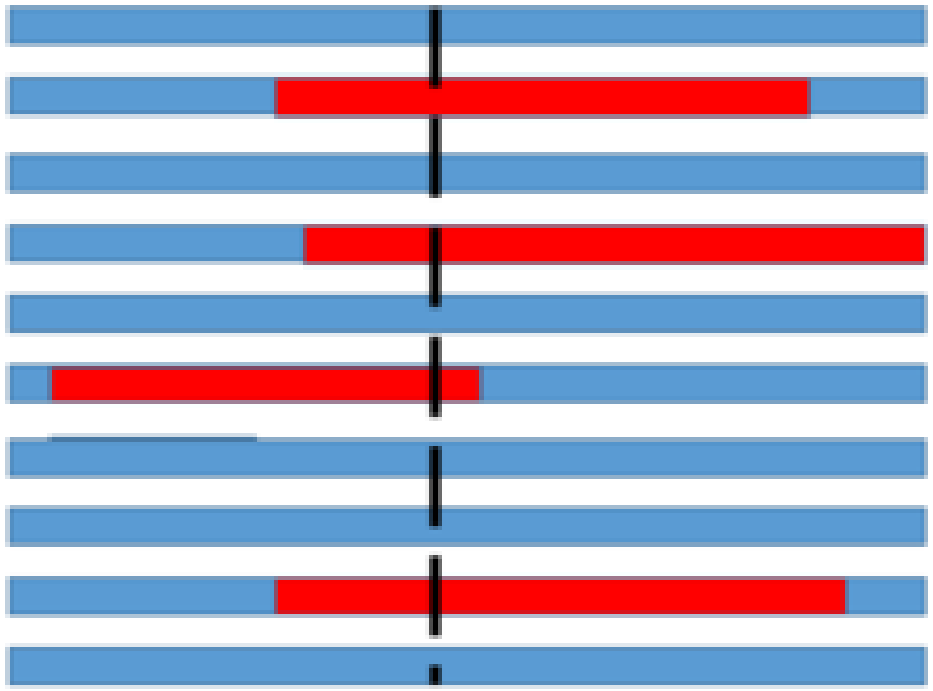
Period Prevalence: 5/10, and if one of the individual died its 5/9

Period Prevalence is taken over a period (1 year for example).

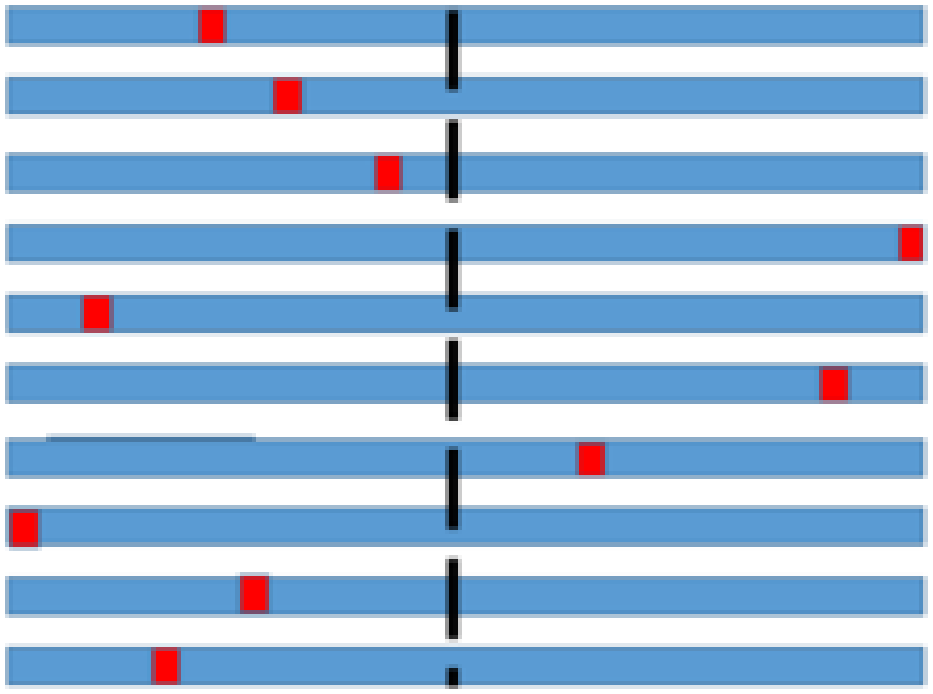
For dynamic populations, take the average of the population size (mid point) , during that time period.

How can prevalence be affected by disease duration?

Disease duration: how long the patient has the disease.



CHRONIC DISEASE: affected 4 in 10, point prevalence = 4 in 10



ACUTE DISEASE: affected 10 in 10, point prevalence = 0 in 10

Notes for the figure above:

Prevalence increases as the disease duration increases and vice versa.

Acute Disease figure: No cases were found (Point Prevalence=0), so prevalence is a good estimate for disease with a long duration of time.

Prevalence is not preferred to be used for infectious diseases, use incidence instead for more accuracy.

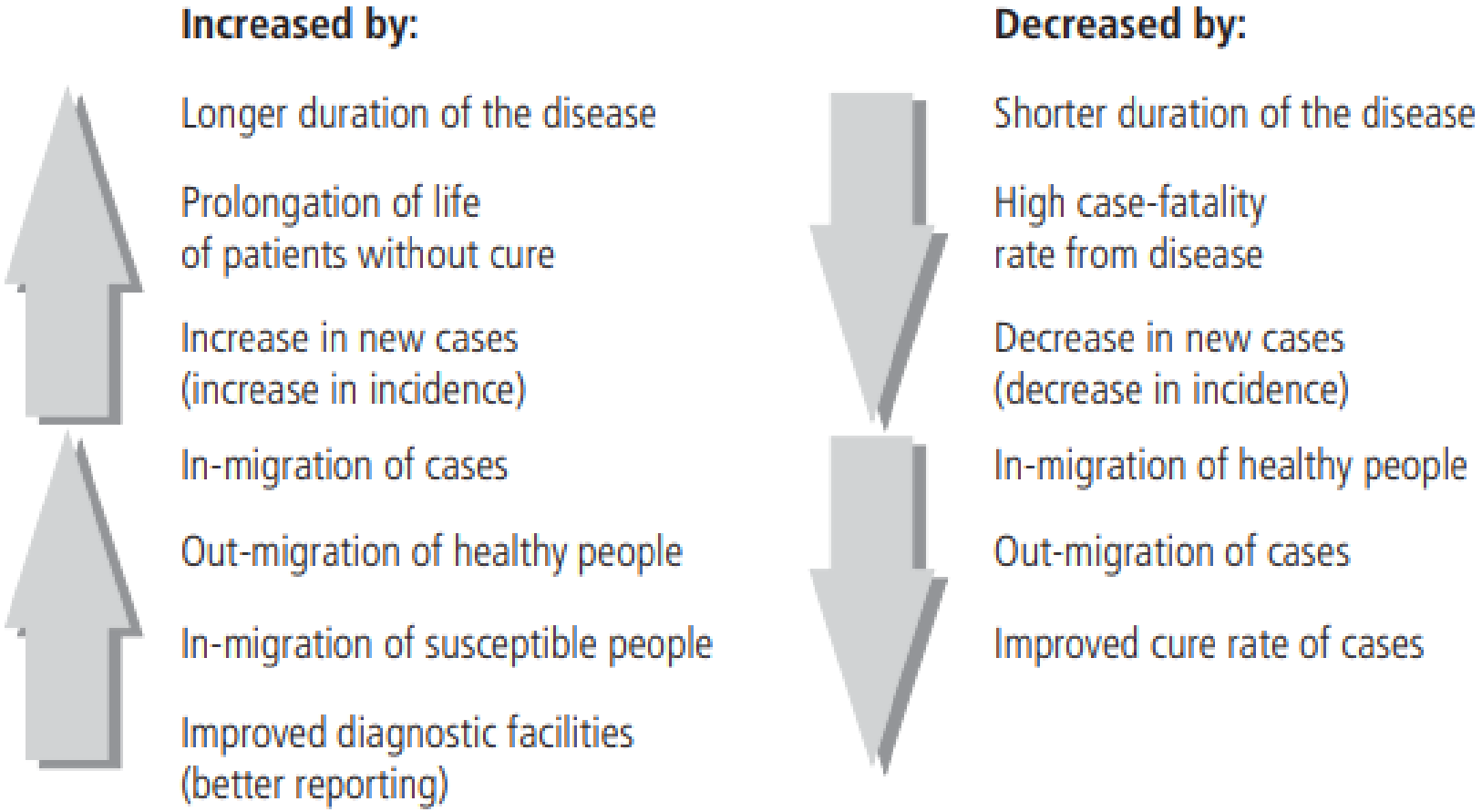
Factors determine prevalence

- The **severity of illness** (if many people who develop a disease die within a short time, its prevalence is decreased);

- The **duration of illness** (if a disease lasts a short time its prevalence is lower than if it lasts a long time). ⟶ increases the numerator

- The **number of new cases** (if many people develop a disease, its prevalence is higher than if few people do so). ⟶ Affects the denominator

Figure 2.2. Factors influencing prevalence



Incidence $\longrightarrow \frac{Number\ of\ new\ cases\ or\ newly\ diagnosed}{Total\ number\ of\ people\ at\ risk} * 10^n$

- **Incidence** refers to the rate at which new events occur in a population.

Incidence *(I)* is calculated as follows:

$$I = \frac{\text{Number of new events in a specified period}}{\text{Number of persons exposed to risk during this period}} (\times 10^n)$$

The numerator strictly refers only to first events of disease. The units of incidence rate must always include a unit of time (cases per 10^n and per day, week, month, year, etc.).

The prevalence is always greater than the incidence in that same period. But if the time period is not matching (different), incidence can be greater than prevalence.

Example: The prevalence of Covid-19 patients in 2019 is less than the incidence of Covid-19 patients in 2020, but the prevalence of Covid-19 patients in 2020 is greater than the incidence of Covid-19 patients in 2020.

Used in Cohort and clinical trials studies.

- **Cumulative incidence** is a simpler measure of the occurrence of a disease or health status. Unlike incidence, it measures the denominator only at the beginning of a study.

The cumulative incidence can be calculated as follows:

$$\textit{Cumulative Incidence} = \frac{\text{Number of people who get a disease during a specified period}}{\text{Number of people free of the disease in the population at risk at the beginning of the period}} (\times 10^{n})$$

The total number of population

Incidence rate person time is used when the population is highly dynamic.

Cross-Sectional survey

Interview Question	Type of Measure
"Do you currently have asthma?"	Point prevalence
"Have you had asthma during the last [*n*] years?"	Period prevalence
"Have you ever had asthma?"	Cumulative incidence

Relationship between incidence and prevalence

Prevalence = incidence x duration

Epidemiologist's bathtub

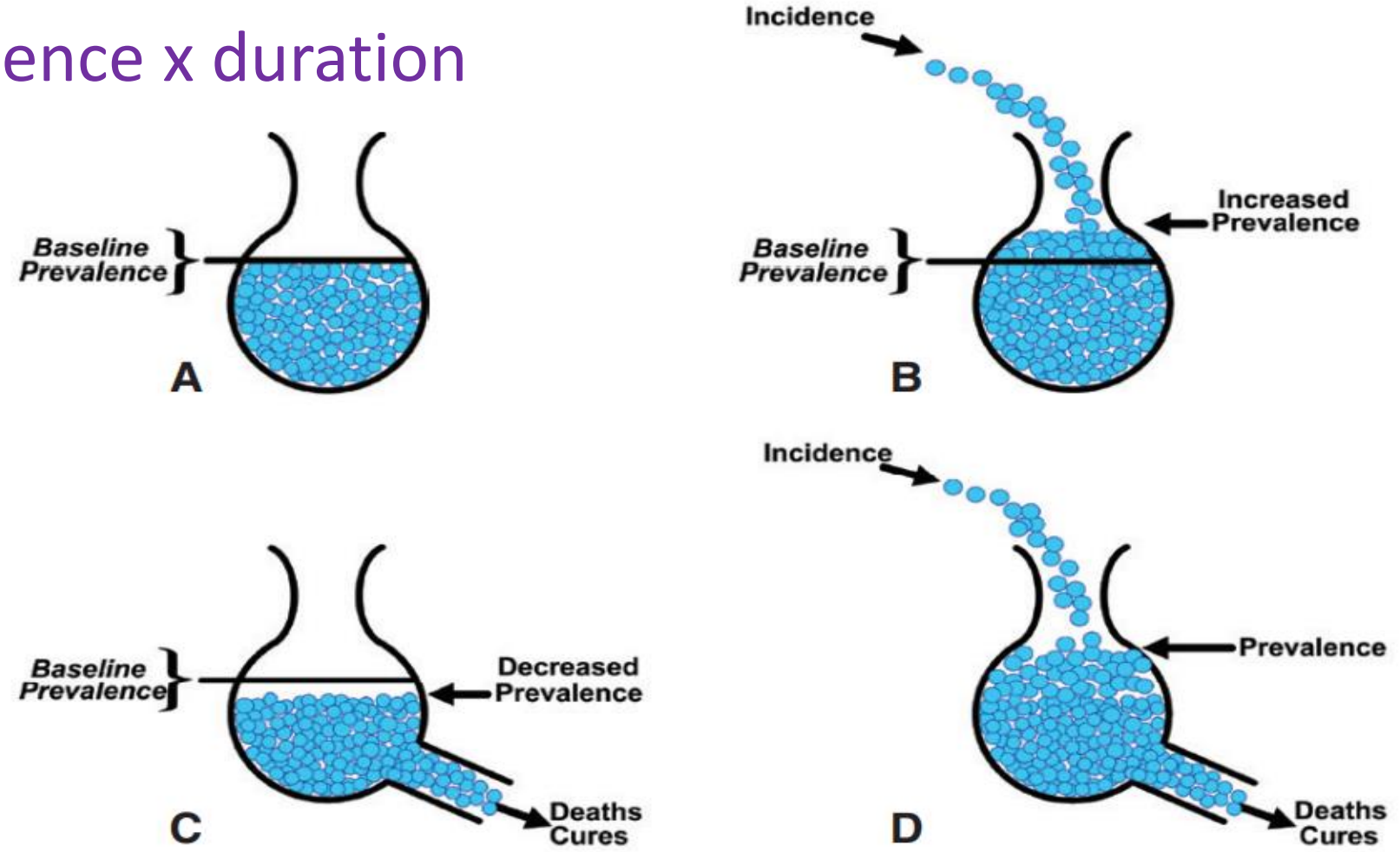


Fig. 3.12 Relationship between incidence and prevalence. (A) Level of prevalence in the population. (B) Increased prevalence resulting from incidence. (C) Decreased prevalence resulting from deaths and/or cures. (D) Overall impact on prevalence of incidence, deaths, and/or cures.

Table 2.2. Differences between incidence and prevalence

	Incidence	Prevalence
Numerator	Number of **new** cases of disease during a specified period of time	Number of **existing** cases of disease at a given point of time
Denominator	Population at risk	Population at risk
Focus	Whether the event is a new case Time of onset of the disease	Presence or absence of a disease Time period is arbitrary; rather a "snapshot" in time
Uses	Expresses the risk of becoming ill The main measure of acute diseases or conditions, but also used for chronic diseases More useful for studies of causation	Estimates the probability of the population being ill at the period of time being studied. Useful in the study of the burden of chronic diseases and implication for health services

Note: If incident cases are not resolved, but continue over time, then they become existing (prevalent) cases. In this sense, prevalence = incidence × duration.

Table 2.2

Incidence can give me a view to the future. Its useful for preventive services (preventive medicine).

Incidence also takes the severity of the disease into account (how severe the disease is), while prevalence won’t.

Questions 6 and 7 use the information below:

Population of the city of Atlantis on March 30, 2012 = 183,000

No. of new active cases of TB occurring between January 1 and June 30, 2012 = 26

No. of active TB cases according to the city register on June 30, 2012 = 264

6 *The incidence rate of active cases of TB for the 6-month period was:*

a. 7 per 100,000 population
b. 14 per 100,000 population
c. 26 per 100,000 population
d. 28 per 100,000 population
e. 130 per 100,000 population

7 *The prevalence rate of active TB as of June 30, 2012, was:*

a. 14 per 100,000 population
b. 130 per 100,000 population
c. 144 per 100,000 population
d. 264 per 100,000 population
e. None of the above

Question 6

Incidence rate of active TB cases $\frac{26}{183000}$ * 100000 = 14.2

(14 per 100000)

Answer is B

Question 7

Prevalence

$\frac{264}{183000}$ * 100000 = 144.26 (144 per 100000)

Answer is C

Measures of mortality (Mortality rate)

- The **death rate** (or crude mortality rate) for all deaths or a specific cause of death.

$$\textit{Crude mortality rate} = \frac{\text{Number of deaths during a specified period}}{\text{Number of persons at risk of dying during the same period}} (\times 10^{n})$$

- **Age-specific death rates**

$$\frac{\text{Total number of deaths occurring in a specific age and sex group of the population in a defined area during a specified period}}{\text{Estimated total population of the same age and sex group of the population in the same area during the same period}} (\times 10^{n})$$

- **Infant mortality**

$$\textit{Infant mortality rate} = \frac{\text{Number of deaths in a year of children less than 1 year of age}}{\text{Number of live births in the same year}} \times 1000$$

Age-specific death rate and infant mortality both are considered as health measures by the WHO.

Example: The death rate in Africa is high compared to the US.

Indications: 1- Bad health care system in Africa.

2- Higher birth rate in the US.

3- Poor disease management in Africa.

Example: Infant mortality rate high in Africa compared to the US.

Indications: 1- poor infant care in Africa

2- poor maternal health care in Africa (Prenatal and postnatal care).

Gives indications about health care related to labor.

- Maternal mortality rate The maternal mortality rate refers to the risk of mothers dying from causes associated with delivering babies, complications of pregnancy or childbirth.

$$\textit{Maternal mortality rate} = \frac{\text{Number of maternal deaths in a given geographic area in a given year}}{\text{Number of live births that occurred among the population of the given geographic area during the same year}} \left(\times 10^{n}\right)$$

Case fatality (CFR)

It is the most important measure that indicates the severity of a disease.

- **Case fatality** is a measure of disease severity and is defined as the proportion of cases with a specified disease or condition who die within a specified time.

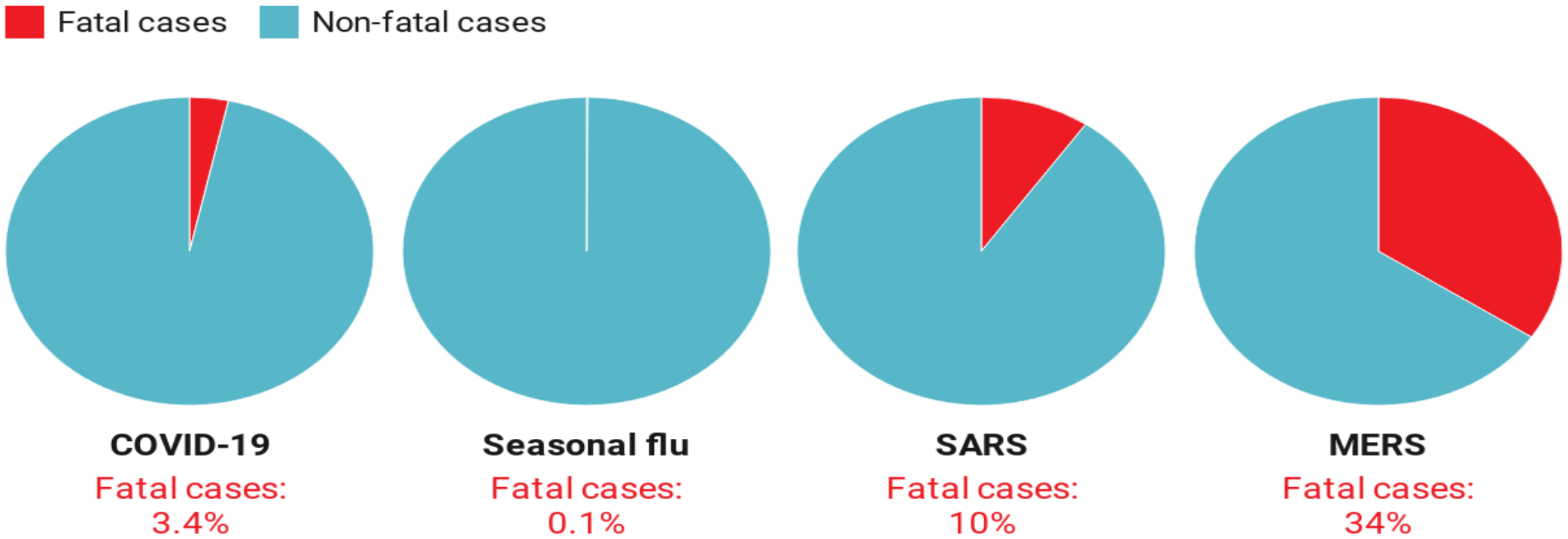
$$Case\ fatality\ (\%) = \frac{\text{Number of deaths from diagnosed cases in a given period}}{\text{Number of diagnosed cases of the disease in the same period}} \times 100$$

Case fatality rate is used for: 1- infectious diseases (mostly)

2- Chronic diseases

Check the denominator ….

COVID-19 looks a lot closer to the season flu than to previous coronavirus outbreaks



COVID-19, SARS, and MERS data are global and total to date. Seasonal flu data are U.S., for the 2018-2019 season.

Chart: Elijah Wolfson for TIME • Source: CDC and WHO • Created with Datawrapper

- **Other population health measures:**
  - Years of potential life lost (PLL) based on the years of life lost through premature death (before an arbitrarily determined age).
  - Healthy life expectancy (HALE) → Takes Life expectancy and Death rate into account.
  - Disability-free life expectancy (DFLE) → The number of years a person lives without being disabled.
  - Quality-adjusted life years (QALYS) → Physical activity, exercise, diet, mental health.
  - Disability-adjusted life years (DALYS) → Takes Death rate and Disability rates into account.

- It combine duration of life with its quality.

Note: The higher the Life expectancy, the more chronic diseases and the higher the disability rates.

Comparing disease occurrence

- The **risk difference** is the difference in rates of occurrence between exposed and unexposed groups in the population.

The risk difference between the incidence rate of stroke in women, who smoke, and the rate of stroke in women who have never smoked, is **31.9 per 100 000** person-years.

Table 2.4. Relationship between cigarette smoking and incidence rate of stroke in a cohort of 118 539 women[13]

Smoking category	Number of cases of stroke	Person-years of observation (over 8 years)	Stroke incidence rate (per 100 000) person-years)
Never smoked	70	395 594	17.7
Ex-smoker	65	232 712	27.9
Smoker	139	280 141	49.6
Total	274	908 447	30.2

Difference

Risk Difference: is the difference between the exposed group to a certain risk factor compared to the unexposed.

Table 2.4

To calculate risk difference:

Firstly, we calculate Incidence rates of strokes for: 1- People who never smoked which is (17.7).

2- People who smoked which is (49.6).

The Risk Difference is 49.6-17.7= 31.9 → This number gives us an idea that there is an association between smoking and strokes.

To understand associations

Gives the risk compared to the population

- **The population attributable risk (PAR)** is the incidence of a disease in a population that is associated with (or attributed to) an exposure to a risk factor.

PAR can be estimated by the formula:

$$PAR = \frac{I_p - I_u}{I_p}$$

where

I_p is the incidence of the disease in the total population and

I_u is the incidence of the disease among the unexposed group.

- This measure is useful for determining the **relative importance of exposures** for the entire population.
- It is the proportion by which the incidence rate of the outcome in the entire population **would be reduced if exposure were eliminated.**

Table 2.4

To calculate Percent reduced if the exposure was eliminated

$$\text{PAR} = \frac{Incidence\ rate\ of\ stroke\ in\ population - Incidence\ rate\ of\ stroke\ among\ non\ smokers}{Incidence\ rate\ of\ stroke\ in\ population} * 100$$

To calculate Percent increased

$$\text{PAR} = \frac{Incidence\ rate\ of\ stroke\ in\ population - Incidence\ rate\ of\ stroke\ among\ smokers}{Incidence\ rate\ of\ stroke\ in\ population} * 100$$

- the attributable risk of smoking for stroke in the smokers is: ((49.6 – 17.7)/49.6) ×100 = 64%.

Table 2.4. Relationship between cigarette smoking and incidence rate of stroke in a cohort of 118 539 women[13]

Smoking category	Number of cases of stroke	Person-years of observation (over 8 years)	Stroke incidence rate (per 100 000) person-years)
Never smoked	70	395 594	17.7
Ex-smoker	65	232 712	27.9
Smoker	139	280 141	49.6
Total	274	908 447	30.2

Difference

Measures of associations

Ratio

- It is the number of (a) observation with specific characteristics divided by the number (b) of the observations without the given characteristics.

Ratio = a / b or a / (n – a)

Usually used to calculate risk ratios like Odds ratio (OR) & relative risk ratio (RR).

- Note that:
- The numerator is not part of the denominator.

Example: Comparing number of smokers relative to nonsmokers

$$\frac{Number\ of\ smokers}{Number\ of\ nonsmokers}$$

My reference population should always be in the Denominator.

Both numerator and denominator are two different groups.

And if Numerator = Denominator, then the ratio equals 1, which means no association.

Relative Risk (RR)

- The **relative risk** (also called the risk ratio) is the ratio of the risk of occurrence of a disease among exposed people to that among the unexposed.

$$\left| \frac{\text{Risk of disease (Cumulative incidence) in exposure group}}{\text{Risk of disease (Cumulative incidence) in non-exposure group}} \right|$$

An RR = 1.00 means that the risk of the event is identical in the exposed and control samples. (no association)

An RR < 1.00 means that the risk is lower in the exposed sample. (Low risk, Denominator > Numerator)

An RR > 1.00 means that the risk is increased in the exposed sample. (High risk, Numerator > Denominator)

Relative Risk:

1- Estimate the incidence rate of diseased among exposed group divided by incidence rate of diseased among unexposed group.

2- Gives a real meaning of the Risk.

3- Used in Cohort studies and clinical trials.

4- Has a Time factor.

Example: Calculate Relative Risk

$$\text{Relative Risk} = \frac{Incidence\ rate\ of\ stroke\ among\ somkers}{Incidence\ rate\ of\ stroke\ among\ nonsmokers}$$

To calculate Incident rates:

Incidence rate of stroke among smokers (A)=

$$\frac{Number\ of\ smokers\ and\ had\ a\ stroke}{Number\ of\ all\ smokers}$$

Incidence rate of stroke among nonsmokers (B)=

$$\frac{Number\ of\ nonsmokers\ and\ had\ a\ stroke}{Number\ of\ all\ of\ nonsmokers}$$

Relative Risk (RR) is $\frac{A}{B}$

Factors	Incidence		Total
	Yes	No	
Exposed group	A	B	A+B
Non-exposed group	C	D	C+D

How many times factor exposure would increase the incidence of an individual:

$$\text{Relative risk} = \frac{\text{Incidence risk among an exposed group}}{\text{Incidence risk among a non-exposed group}} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

Relative risk larger than 1 represents that risks have increased due to factor exposure.

The value obtained by subtracting 1 from the relative risk is an excess relative risk, showing an increased amount of risks.

RR for preventive measure, example Physical activity RR = 0.5, which means that people who exercise are 0.5 less likely to develop disease compared to people who don't exercise.

How many times more likely are exposed persons to become disease, relative to non-exposed persons?

- **<u>Example an RR = 1.9 means, people whom got exposed are 1.9 likely to develop the disease compared to those didn't exposed.</u>** (Interpretation of Relative risk).

Odds Ratio (OR)

Gives an approximation of the risk

Used for studies that lack time follow up like cross-sectional studies and case control studies.

- The odds ratio is a measure of association used when we can calculate incidence rate.

- **Odds ratio** refer to the odds of people with the disease (cases) may got exposed.

- **Similar to RR:**
- OR = 1 , no association or relationship.
- An OR < 1.00 means that the risk/odds is lower in the exposed sample.
- An OR > 1.00 means that the risk/odds is increased in the exposed sample.

Calculating the Odds Ratio (OR)

	Disease (Case)	No Disease (Control)
Exposed	A	B
Unexposed	C	D

$$OR = \frac{\text{Odds that a case was exposed } (A/C)}{\text{Odds that a control was exposed } (B/D)} = \frac{AD}{BC}$$

How many times more likely are persons with the disease being exposed compared to persons don't have the disease?

- Example an OR = 0.5 means, the odds of people with the disease got exposed are 0.5 less likely compared to those don't have the disease.

- Example to calculate Odds Ratio:

$$\frac{Number\ of\ smokers\ who\ had\ stroke}{Number\ of\ nonsmokers\ who\ had\ stroke} \div \frac{Number\ of\ smokers\ who\ never\ had\ stroke}{Number\ of\ non\ smokers\ who\ never\ had\ stroke}$$

In Odds ratio, the numerator is always the diseased group.

Exposure	Outcome		
	Migraine	No migraine	Total
Sumatriptan	27	273	300
Placebo	108	212	320
Total	135	485	620

To calculate Relative risk

$$\frac{Incidence\ of\ migraine\ amog\ people\ who\ take\ treatment}{Incidence\ of\ migraine\ amone\ people\ who\ don't\ take\ treatment}$$

$\frac{27}{300} \div \frac{108}{320}$ = 0.27 ⟶ The medication decreased migraine events by 0.27

To calculate Odds ratio

$\frac{27}{108} \div \frac{273}{212}$ = 0.194 ⟶ The odds of migraine decreased by 0.194 after using Sumatriptan in comparison to Placebo.

Note: Usually RR is greater than OR, except in one condition which is rare diseases, here OR will be close to RR.

Effectiveness of sumatriptan as a treatment for migraine:

Risk difference (RD) = 27/300 - 108/320 = - 0.248

Risk Ratio (RR) = 27/300 ÷ 108/320 = 0.267

Odds ratio (OR) = (27x 212) ÷ (273 x 108) = 0.194

Questions