
MEDICAL RESEARCH
MODIFIED NO. 11
كتابة: دكتور ٢٠٢١
تدقيق: ميس مصطفى
الدكتور: منير أبو هلالة

Overview of study design part 1: Analytical studies (cohort studies)

Dr Munir Abu-Helalah

MD,MPH,PHD

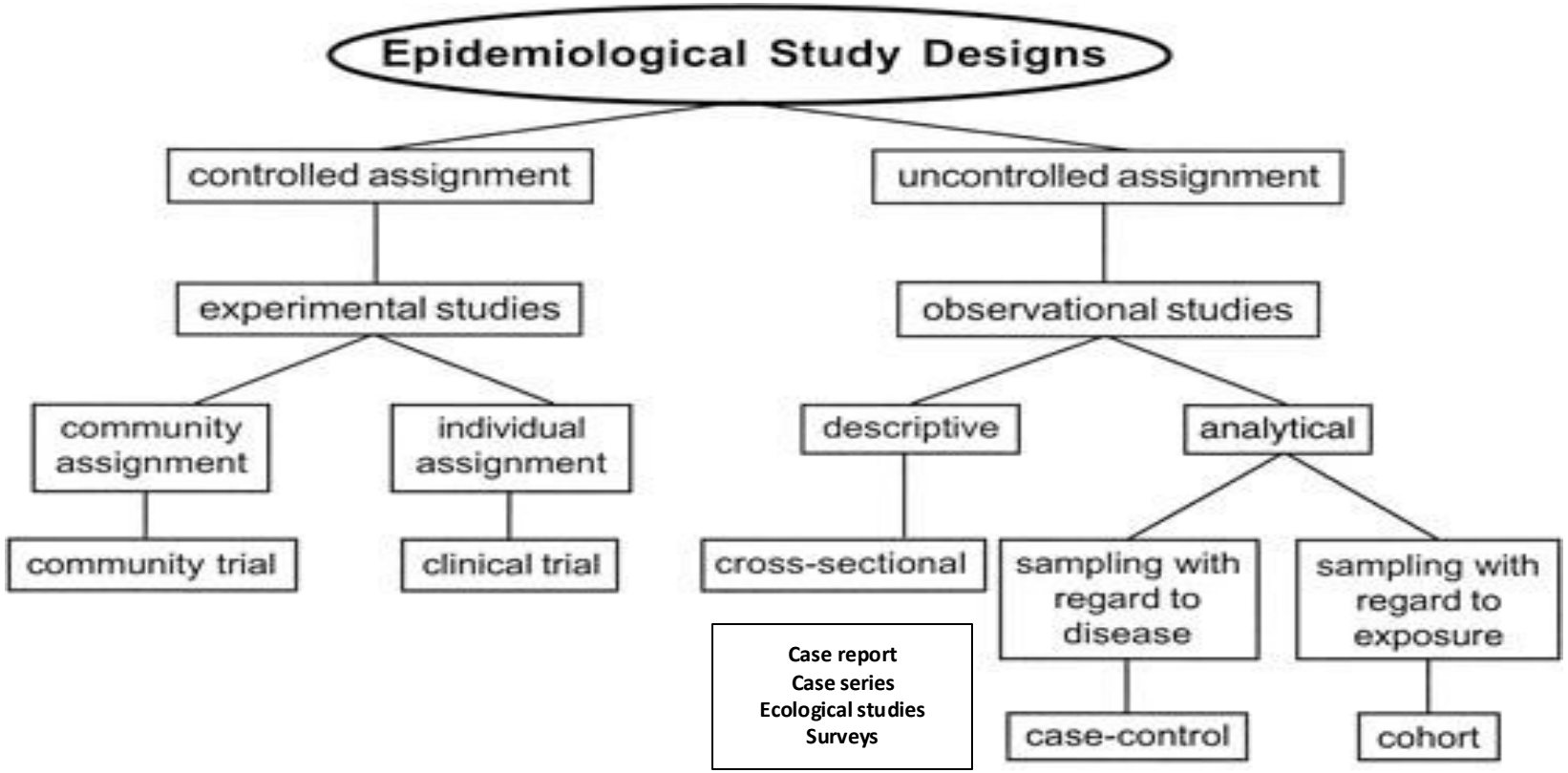
Associate Professor of Epidemiology and Preventive Medicine

This week, we cover analytical studies. We previously covered observational, descriptive studies such as cross-sectional studies, surveys, ecological studies, case reports, and case series. These studies help us generate hypotheses about potential risk factors.

Color code

- Slides
- Doctor
- Additional info
- Important

Note: This Modified was written by Doctor 2021, but it has been paraphrased for improved wording. Good luck !!

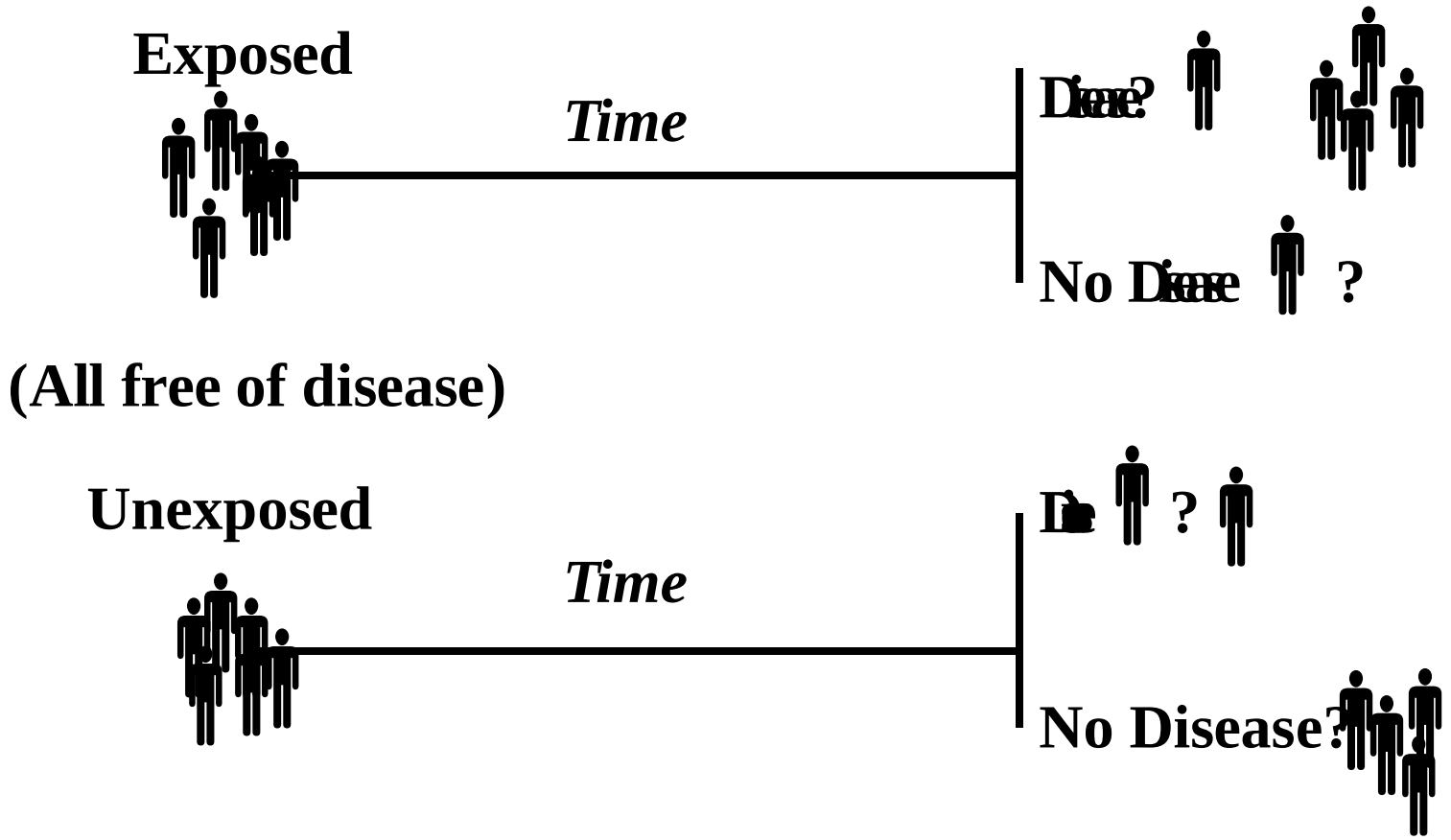


Source: Waning B, Montagne M: *Pharmacoepidemiology: Principles and Practice*: http://www.accesspharmacy.com

For instance, we observed that 17 out of 20 patients in a case series had human papillomavirus (HPV) positivity, and in the cross-sectional study, the prevalence of HPV positivity was higher among these patients than among the general population or patients who were presenting with other conditions.
We discovered that these potential risk factors (observations) are highly prevalent among patients with various illnesses. We also found that Helicobacter pylori and stomach cancer may be risk factors for stomach cancer, and smoking may be a risk factor for diabetes, hypertension, hypothyroidism, or cancer. Afterward, we must conduct analytical studies to prove or disprove these findings.

Cohort studies and case-control studies are the two categories of analytical studies that we have discussed in brief during the past few weeks.
Cohort studies are a classic example of analytical research; at baseline, we have two groups—one exposed and one unexposed—that should be similar in age and gender and free of the disease under investigation. We then follow these groups for ten or twenty years to examine the prevalence of various illnesses.

Cohort studies



For instance, if we have two groups—Group A, which consists of smokers, and Group B, which includes nonsmokers—and we track the incidence of health issues such as ischemic heart disease, diabetes, cancer, and various respiratory illnesses, we might observe that the incidence rate of ischemic heart disease in smokers is 100 cases per 10,000 people, while for nonsmokers, it is 20 cases per 10,000. This implies a relative risk of 5 (100/10,000 divided by 20/10,000), indicating that smokers have a five times higher risk of developing the disease compared to nonsmokers.

This table is additional

	Developed Ischemic Heart Disease	Did Not Develop Ischemic Heart Disease	Total
Smokers	100	9,900	10,000
Nonsmokers	20	9,980	10,000
Total	120	19,880	20,000

The relative risk (RR) calculation:

$$RR = \frac{\text{Incidence in Smokers}}{\text{Incidence in Nonsmokers}} = \frac{100/10,000}{20/10,000} = 5$$

So, smokers have a five times higher risk of developing ischemic heart disease than nonsmokers.

Cohort studies generally involve comparing two groups: an exposed group, consisting of individuals with a specific risk factor, and an unexposed group, made up of individuals without that risk factor. These groups are often matched for variables like age and gender to ensure comparability. For instance, if the exposed group consists of smokers aged 20 to 40 from a particular geographical area, the unexposed group should also include individuals aged 20 to 40 from the same area, with no significant baseline differences in age or other factors that could influence disease incidence.

Cohort studies can investigate a variety of risk factors. Examples include comparing smokers to nonsmokers, individuals positive for HPV to those negative, people taking aspirin for non-trial-related reasons (e.g., to reduce heart disease risk or prevent recurrent myocardial infarction) to those not taking aspirin, or individuals following a high-fiber diet versus those who are not. Other comparisons might involve physically active versus inactive individuals, occupational risk factors, or various lifestyle behaviors.

A key strength of cohort studies is their ability to examine multiple health outcomes associated with a single risk factor, making them highly versatile for studying the relationships between risk factors and diseases.

In cohort studies, a critical requirement is that neither the exposed nor the unexposed groups have the disease of interest at the baseline. This ensures that we can accurately compare the relative risk of developing the disease based on exposure.

For example, if the goal is to study the effect of smoking on the development of type 2 diabetes, two steps must be taken at the baseline:
1. Ask participants in both the exposed (smokers) and unexposed (nonsmokers) groups about any history of diabetes, impaired glucose function, or pre-diabetes. Individuals with these conditions should be excluded from the study.
2. Conduct glucose testing, including fasting glucose levels and HbA1c, to confirm normal glucose function in all participants.

Similarly, for a study examining the relationship between smoking and hypothyroidism, participants should be screened for any history of hypothyroidism at the baseline, and TSH levels should be measured to ensure normal thyroid function.

If the study focuses on breast cancer risk factors, participants should be screened appropriately based on their age: younger individuals (under 40) should undergo MRI, while older participants (40 and above) should have a mammogram.

In all cases, it is essential that participants are free from the disease of interest at the start of the study. Follow-up is then conducted to monitor the development of the disease over time.

Cohort studies are beneficial because they allow us to calculate the incidence of a condition during the follow-up period, which is essential for determining the relative risk.

To calculate incidence:
If 100 patients develop hypothyroidism over 10 years, the annual incidence is 10 cases per year. For a population of 10,000, this translates to 10 cases per 10,000 per year. To standardize the measure, we scale it to a population of 100,000, resulting in 100 cases per 100,000 annually.

For shorter follow-up periods, the incidence must be adjusted to reflect annual rates. For instance, if 50 cases are observed in 6 months, doubling this number gives an annual incidence of 100 cases. Conversely, for longer follow-ups, such as two years, dividing the total cases (e.g., 100 cases) by two gives an annual incidence of 50 per the relevant population size.

One major advantage of cohort studies is their ability to establish temporal relationships. This means we can be confident that the exposure or risk factor occurred before the development of the disease. This is achieved by thoroughly assessing participants at baseline, including asking about their medical history and conducting tests to confirm they were disease-free at the start of the study. This clear sequencing strengthens the validity of findings related to cause and effect.

Cohort (or follow-up) studies

- **Are studies in which people are identified and grouped with respect to whether or not they have been exposed to a specific factor.** (occupation, lifestyle risk factors or taking certain medications).

- **The groups are followed up over time to determine whether the incidence of a particular disease is any greater (or less) in the exposed group than in the non-exposed group.**

- **The starting point is the risk factor!** (in rare diseases, explained later)

A cohort study involves observing and comparing groups over time without any intervention by the investigators. For example, if researchers want to examine the impact of daily multivitamin intake on various diseases, they would conduct a cohort study by comparing individuals who already take multivitamins regularly with those who do not. However, if the researchers were to distribute multivitamins to participants, it would instead be a clinical trial.

Similarly, to investigate whether aspirin use reduces the incidence of colorectal cancer, researchers could review hospital records to identify individuals who already take aspirin regularly and compare their outcomes with those who do not. This would qualify as a cohort study. Conversely, assigning participants to receive aspirin or a placebo would make it a clinical trial.

In cohort studies, like other observational designs such as case-control studies, researchers do not intervene in any way, such as adjusting doses or frequencies. Instead, they follow the groups over time to assess whether a specific exposure increases or decreases the incidence of a disease. For instance, the incidence of disease may be higher among individuals exposed to risk factors like smoking, while it may be lower in those who take protective measures, such as using aspirin.

Cohort study examples:

- Life expectancy of cerebral palsy children
- Fine needle breast biopsy and breast cancer
- Aspirin intake and colorectal cancer

Cohort study:
Primary purposes

- **Descriptive (measures of frequency)**

– To describe the incidence rates of an outcome over time, or to describe the natural history of disease

- **Analytic (measures of association)**

– To analyze associations between the rates of the outcomes and risk factors or predictive factors

COHORT STUDY DESIGN

- This design is the best observational one for establishing cause-effect relationships.
- Prevention and intervention measures can be tested and affirmed or rejected.
- Cohort studies consider seasonal variation, fluctuations, or other changes over a longer period.

To calculate the incidence of conditions like gastroenteritis, influenza infections, or myocardial infarctions (MI) over a one-year period, a cohort study is necessary. Cross-sectional studies are not suitable for such purposes because they capture data at a single point in time and cannot effectively account for short periods (e.g., 5–6 months) or seasonal variations in disease patterns. Cohort studies, by following individuals over time, allow for a more accurate assessment of these fluctuations and the incidence of such conditions.

- Objective measures of exposure, such as biological markers, are preferred over subjective measures. Cohort studies are particularly useful for assessing biological markers, such as smoking, pre-diabetes, and type 2 diabetes. These markers are more reliable because they remain consistent over time, unlike subjective measures, which can fluctuate. For example, tracking subjects with depression (a subjective measure) in a follow-up study can be challenging due to the variability in their condition over time. This is why cohort studies typically focus on biological markers, as they provide a stable and objective basis for analysis.

COHORT STUDY DESIGN
Strengths

- **We can measure incidence of disease in exposed and unexposed groups**
- **Can get a temporal (time related) sequence between exposure and outcome as all individuals must be free of disease at the beginning of the study.**
- **Good for looking at effects of rare exposures.**
- **Allows for examination of multiple effects/diseases of a single exposure.** (Ensure that the subjects are free of the disease of interest at baseline)
- **Not open to bias as much as other types of study** (in case-control studies we have interviewer, selection, recall bias, etc.)
- **Direct calculation of the risk ratio or relative risk is possible.** (In case-control studies, relative risk cannot be calculated because we only know the number of cases with the disease, not the total population, meaning the denominator for risk calculation is missing. Incidence cannot be determined either, so we use the odds ratio instead. Relative risk compares the risk of an event between exposed and non-exposed groups.)
- **Provide information on multiple exposures**

.

COHORT STUDY DESIGN

Limitations:

- Not efficient for rare diseases (A disease is considered rare if its incidence is less than 1 per 10,000, and calculating it requires a very large sample size)
- Can be expensive and time-consuming
- Large sample
- Drop-out biases (With a large sample size, some participants may relocate or be lost during follow-ups, and risk factors can change over time (e.g., smokers quit or nonsmokers start smoking). Incidence density can adjust for this, but follow-up studies still have limitations.)

  If the study goes over many years, it can cause considerable loss to follow-up. This can 'dilute' results or lead to bias, and therefore the validity of results can be seriously affected

- Locating subjects, developing tracking systems, and setting up examination and testing processes can be difficult.
- Changes over time in diagnostic methods, exposures, or study population may lead to biased results.

Cohort study: Example

Hypertension as a risk factor for spontaneous intracerebral hemorrhage

In study risk factors, we start with what is rare!

- Rare disease: we conduct a case-control study starting with cases
- Rare risk factor: we conduct a cohort study starting with rare risk factors

For rare diseases, case-control studies are typically used, while for rare risk factors, cohort studies are more appropriate. This is because identifying rare occurrences can be challenging in the context of certain study designs.

For example:
If studying the effects of radiotherapy in children treated for childhood malignancies on the likelihood of developing adult cancers, or the impact of fine-needle aspiration for benign breast tumors on the risk of later breast cancer, case-control studies might struggle to capture such rare exposures. Interviewing patients with specific illnesses may not yield sufficient data on these uncommon risk factors.

Instead, when investigating rare risk factors, start by focusing on the risk factor itself. For instance:
- To study the effects of childhood radiotherapy, examine hospital records of patients treated for childhood malignancies. Track these children over time and compare outcomes with those who didn't receive radiotherapy or had no malignancy.
- To explore the link between fine-needle aspiration and breast cancer, review histopathology records for women who underwent benign needle biopsies. Compare breast cancer incidence in these women with others who visited the same hospital but did not undergo needle biopsies.

By using cohort studies in such scenarios, you can effectively track and analyze the outcomes of rare exposures.

For rare diseases, case-control studies are preferred because the low incidence makes it impractical to use other methods. Identifying enough cases in a population would require an exceptionally large sample size. For example, if a disease occurs at a rate of 1 in 100,000, a cohort study would need to follow 100,000 individuals for 10 years just to observe 10 cases. This approach is neither feasible nor efficient.

The key principle is to start with the subjects who already have the condition or risk factor:

- For rare diseases, use case-control studies, selecting individuals who already have the disease at the study's outset.
- For rare risk factors, use cohort studies, focusing on individuals who already possess the risk factor and tracking their outcomes over time.

This approach ensures that the study design aligns with the rarity of the condition or exposure, optimizing efficiency and feasibility.

In cross-sectional studies, rare diseases cannot be effectively studied to determine their magnitude because identifying a sufficient number of cases would require an extremely large sample size.

To assess the burden or magnitude of a rare disease or one with a short duration, cohort studies are more appropriate. For instance, to study the burden of congenital heart diseases in Jordan, a cohort study could involve tracking pregnant women over two years to calculate the incidence of congenital heart disease in their babies. However, if the goal is to investigate the risk factors for congenital heart diseases, a case-control study would be more suitable.

In medical literature, discussions of common diseases typically focus on prevalence, while rare diseases emphasize incidence. This is because assessing the burden of rare illnesses relies on calculating incidence, which is best achieved through cohort studies.

Physical Activity and Incident Cognitive Impairment in Elderly Persons

ARCH INTERN MED/VOL 170 (NO. 2), JAN 25, 2010

Background: Data regarding the relationship between physical activity and cognitive impairment are limited and controversial. We examined whether physical activity is associated with incident cognitive impairment during follow-up.

Methods: As part of a community-based prospective cohort study in southern Bavaria, Germany, 3903 participants older than 55 years were enrolled between 2001 and 2003 and followed up for 2 years. Physical activity (classified as no activity, moderate activity [<3 times/wk], and high activity [≥3 times/wk]), cognitive function (assessed by the 6-Item Cognitive Impairment Test), and potential confounders were evaluated. The main outcome measure was incident cognitive impairment after 2 years of follow-up.

Cohort study

Physical activity	**Cognitive impairment** Yes	No	Total
Moderate	10	990	1000
None	100	900	1000
Total	110	1880	2000

Risk of outcome in exposed (not active) = 100/1000 = 10%

Risk of outcome in non-exposed (active)=10/1000 =1%

Relative risk 10%/1%=10 =

This cohort study examines the effect of physical activity on cognitive impairment. It included 3,903 participants over 55, followed for two years. Physical activity levels were categorized as none, moderate, or high. Cognitive impairment was assessed to determine the impact of activity. The exposed group included those with no activity (risk factor), while those with moderate activity lacked the risk factor.

Risk of outcome in exposed (not active)= 100/1000 = 10%

This means that cognitive impairment incidence for the exposed group is 10%.

Risk of outcome in non-exposed (active)= 10/1000 =1%

Relative risk 10%/1%= 10

This means that subjects who are physically inactive are 10 times at high risk of having cognitive impairment compared with physically active subjects.

Measurement of risk

$$Risk\ (R) = \frac{\text{No of people becoming ill during the period of observation}}{\text{No of people exposed at the beginning of the period}}$$

It is proportion (0 - 1)

Incidence is measured over time; 100 cases in two years equal 50 per year, while 50 cases in six months equal 100 per year.

Hazards and the risks

- Hazards and the risks associated with them are everywhere, but when known measures can be taken to minimise or eliminate risk. When we go up or down stairs it is possible that we might fall, but the likelihood is that we will not.
- Stairs are a hazard, the likelihood of injury is known as the risk. The latter is often expressed as a fraction like 1 in 100 or 1 in a million.

For example, broken stairs (hazards) increase the risk of injury compared to intact stairs.

Measuring the association between risk factor and diseases

Relative risk

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

- **RR=1**

  There is no association between exposure and disease.

- **RR>1**

  Exposure is associated with an *increase* of the frequency of the disease.

- **RR<1**

  Exposure is associated with a *decrease* of the frequency of the disease. (Preventive factor)

- **The value of the RR reflects the magnitude of the association between exposure and disease.** (The higher the relative risk, the stronger the association between the risk factor and the disease

- **RR=5 means that the probability to develop the disease in the exposed is 5 times the probability to develop it in the non exposed**

Calculation of the relative risk

Cohort study

	Disease Present	Disease absent	
Exposure Present	a	b	a+b
Exposure absent	c	d	c+d
Total	a+c	b+d	a+b+c+d

	Disease Present	Disease absent	
Exposure Present	a	b	a+b
Exposure absent	c	d	c+d
Total	a+c	b+d	a+b+c+d

Risk in the exposed=(a)/(a+b)

Risk in the non exposed=(c)/(c+d)

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

Here's another example of risk rates: women using contraceptives are 1.4 times more likely to develop bacteriuria. It's crucial to examine the confidence interval. In t-tests, the reference point is zero. If the confidence interval ranges from -1 to 3, it includes zero and is not significant. However, in relative risk or other ratios, the reference point is 1. For a result to be significant, the confidence interval must not cross 1. If the interval is higher than 1 (e.g., 1.2 to 1.8), the factor is significant. However, if it includes or is less than 1 (e.g., 0.9 to 1.6), it is not significant.

Example

Data from a cohort study of oral contraceptive (OC) use and bacteriuria among women aged 16-49 years

	Bacteriuria		
	Yes	No	Total
OC use			
Yes	27	455	482
No	77	1831	1908
Total	104	2286	2390

Data from D. A. Evans et al., Oral contraceptives and bacteriuria in a community-based study. *N. Engl. J. Med.* 299:536, 1978.

$$\text{Relative Risk}\,(RR) = \frac{27/482}{77/1908} = 1.4$$

Example

- **Rate of malaria among illiterate is 8/1000**
- **Rate of malaria among literate is 4/1000**
- **Rate ratio is 2**
- **This means that those who are illiterate have twice the rate of malaria than those who are literate**
- **Literacy is a marker rather than a causal risk**

Preventive fraction

If the exposure is preventive $I_{exposed} < I_{unexposed}$

$$PF = \frac{I_{\text{unexposed}} - I_{\text{exposed}}}{I_{\text{unexposed}}}$$

If the relative risk (RR) is greater than 1, it indicates a risk factor. If RR is less than 1, it represents a preventive factor. The preventive fraction is calculated as the incidence in the unexposed group minus the incidence in the exposed group, divided by the incidence in the unexposed group.

Example

Ischaemic heart disease (IHD) as a disease outcome and exercise as a preventative exposure.

	IHD risk
Exercise	2/100
No exercise	8/100

$$PF = \frac{8/100 - 2/100}{8/100} = 0.75$$

0.75 as a proportion can also be expressed as percentage, 75%. We can say that 75% of the cases of IHD in people who do not exercise could be prevented by exercise.

To calculate relative risk, using no exercise as the risk factor (exposed group), it is 8% ÷ 2% = 4. This means lack of physical activity increases the risk of developing IHD by 4 times compared to physically active individuals.

To assess the impact of exercise in preventing ischemic heart disease, we use the preventive fraction. This is calculated by subtracting the incidence in the exposed group (8%) from the incidence in the unexposed group (2%), and then dividing by the incidence in the exposed group (8%). This gives 0.75, or 75%. This means that 75% of ischemic heart disease cases in individuals who do not exercise could be prevented by exercising.

However, we cannot claim that the entire 8% of ischemic heart disease cases in non-exercisers are solely due to lack of exercise, as there are other contributing risk factors. The actual difference (risk of developing IHD from not exercising) is 6%, not 8%. While lack of exercise is an important risk factor for ischemic heart disease in this group, it is not the only factor. Other risk factors also contribute to the 8% observed in the table.

Additionally, you can observe that the incidence of ischemic heart disease with exercise is not zero, but 2%. Therefore, we do not compare the non-exercising group with zero; we compare it to the 2% incidence. This is why we say the difference is 6%, not 8%. The percentage of developing IHD due to not exercising and other risk factors is calculated as the total incidence (8%) minus the incidence in the exercise group (2%), resulting in 6%.

Design of cohort studies

1. Research question must be clear
2. Set the sample size
3. Set the follow-up period (immediate, short term and long term)
4. Specify study group Sample must be representative of the population you are studying
5. All participants should be free of the outcome (disease) at the beginning of the study
6. Must be able to get correct information about exposure status easily
7. Measure the outcome
8. Comparison group must be as similar as possible to exposed group
9. Put measures in place to reduce loss to follow up if possible

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Influenced by a variety of factors including:

1. Type of exposure being investigated
2. The frequency of the exposure in the population
3. The accessibility of subjects.

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Exposed and unexposed subjects must be free of the outcome of interest at the start of the study and equally susceptible to developing the outcome during the course of the study.
- If some subjects already have the outcome (e.g., disease) at the onset, then the temporal relationship between exposure and outcome becomes obscured. (Subjects should be free from the disease at the baseline to conform the temporal relationship).

We also need participants who are cooperative and stay in the study long-term, not individuals who only work in a specific area for a limited time and then leave, as this could affect a 10-year follow-up. Additionally, we must establish inclusion criteria to determine who should be part of the study, based on factors like gender and specific behaviors, such as smokers who smoke at least once a day. For example, individuals who smoke every other day or ex-smokers would not be included in the study.

The level of surveillance should be consistent between the exposed and unexposed groups. This means both groups must meet the inclusion criteria and then be classified as smokers and non-smokers. It is also important that the control group does not include ex-smokers, as they may also be at risk for certain illnesses. Both groups should undergo the same frequency of examinations, such as every six months for 10 years. Additionally, we must ensure that, at baseline, the groups are comparable, with no significant differences in age, gender, or other risk factors, which are referred to as confounding factors.

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Each subject must rigidly satisfy the criteria for inclusion in the cohort study, and he or she should not be excluded from subsequent analysis because of any change in exposure status during follow-up.
- The degree of surveillance should be similar in exposed and unexposed groups.
- Frequency of examination and duration of follow-up depend on the type of exposure and the outcome under investigation.

COHORT STUDY DESIGN

Selection of subjects for a cohort study

- Both groups should be accessible and available for follow-up.
- Multiple comparison groups for exposed subjects chosen in different ways may reinforce the validity of findings.

Types of cohorts

- **Birth cohort : all individuals in a certain geographic area born in the same period (usually a year)**

In a birth cohort study, we follow subjects who share the same birth year and place, allowing us to study multiple risk factors for various illnesses, rather than focusing on just one, as in an exposure cohort. The key aspect of a birth cohort study is that, at baseline, we need to gather a comprehensive list of all relevant information about the subjects, including risk factors and illnesses to be studied. For example, a study could examine physical activity, fat intake (dietary factors), and smoking (socio-demographic factors) in individuals born in the year 2000, or in people aged 20 living in the same city, and follow them for 20 years.

- **Inception cohort: all individuals assembled at a given point based on some factor, e.g. where they live or work**

For example, we could follow up workers at the same facility or medical students from the same university to examine the risk factors and the development of illnesses.

- **Exposure cohort: individuals assembled as a group based on some common exposure**

the classical type of cohort study, it can only study one risk factor for many illnesses.

  - **e.g. smokers**
  - **e.g. radiation**

Healthy worker effect

A phenomenon of workers usually exhibiting overall death rates lower than those of the general population due to the fact that the severely ill and disabled are ordinarily excluded from employment.

It depends on the workplace. For example, in institutions like JU or JUH, after 20 years of follow-up, individuals with severe illnesses or disabilities might have changed jobs, lost their jobs, or left, which means the sample would no longer be representative of the entire population. This is why it's important to have samples from the general population, not just from a group of healthy workers. In the 70s and 80s, many cohort studies were conducted in workplaces, but this approach is no longer encouraged.

COHORT STUDY DESIGN

- Measurement of exposures should be based on intensity, duration, regularity, and variability. For example, in a study on physical exercise or smoking, we need to gather details such as the duration, frequency, and other related factors.
- Some exposures are acute, one-time episodes never repeated in a subject's lifetime. For example, in a study involving people who received certain medications, such as children who underwent radiation therapy for childhood malignancies, we followed them up even though they had the same exposure with no change over time.
- Other exposures are long-term, such as cigarette smoking use of oral contraceptives, or exercise.
- Exposures may also be intermittent.

COHORT STUDY DESIGN

Retrospective cohorts

- Uses information on prior exposure and disease status.
- All of the events in the study have occurred and conclusions can be drawn more rapidly.
- Costs can be lower
- May be the only feasible one for studying effects from exposures that no longer occur, such as discontinued medical treatments.
- The main disadvantage of a retrospective cohort study is that the investigator must rely on existing records or subject recall. There may be missing or incomplete data at baseline and throughout the study period. Retrospective studies have the limitation of potentially incomplete data, so in such cases, it is better to conduct the study prospectively, as in a classical cohort study.

Retrospective cohort

- If we already have data from previous studies, it should not be done prospectively, as it would be time-consuming. Instead, we can conduct the study quickly as a retrospective study. To do this, we can access records from the family medicine department or use general practice records. We would then review the records for a specific duration and divide the files at baseline into two groups: medical notes of smokers and medical notes of non-smokers.

- Smoking and type II DM
- We start from the year 2002 and follow up for 20 years until 2022, with 5000 smokers, 5000 non-smokers
- In the year 2002 we split the files into: Medical notes of smokers versus medical notes for non-smokers
- Both groups should not have diabetes or impaired glucose profile at baseline
  so in retrospective studies, we should also make sure that at the baseline subjects don't have the disease of interest, then we look through the files and see over 20 years who had developed type two diabetes
- Then, we measure the incidence of Type II DM in the smoking and no-smoking groups.
- The follow-up was completed in the past, therefore, we call it a retrospective cohort study.

Ambidirectional Cohort

- **Data collected both retrospectively and prospectively on the same cohort to study short and long-term effect of exposure**
- **If medical notes in the previous example were incomplete in 2002 but more complete and accurate data are available since 2015.**
- **From the year 2015 until date, the follow-up is in the past, if we continue for an additional 12 years. This means a combination of retrospective and prospective data.**

Imagine we are in 2015 and want to conduct a retrospective study over a 20-year period. While reviewing the files, we realize that the data from 2002 and earlier is insufficient for the study. We can't rely solely on the data from 2002 to 2015 because we need a full 20 years of data. The solution would be to combine the available retrospective data with prospective data collection. By 2023, we would have another 8 years of data, completing the 20-year duration. This approach is called an ambidirectional cohort study. These studies are especially useful in developing countries, where past records may be incomplete. If there isn’t enough data from a specific year for a retrospective study, you can start with the complete records you have and then follow up prospectively, filling in the gaps with future data.

COHORT STUDY DESIGN

Loss during follow-up

- Following subjects over a long period of time can lead to a variety of problems.
- Dropouts and losses of subjects to follow-up are major problems in cohort studies.
- Subjects may move away or leave the study for other reasons, including deaths from other causes than the disease under investigation.
- If losses to follow-up are significant during the study, then the validity of the results can be seriously affected.

The main limitation of cohort studies is the difficulty in studying rare diseases, as a large sample size and long duration are required to gather enough cases. Another challenge is the loss of subjects during follow-up. Participants may lose interest in the study, move to a different location, or die from causes unrelated to the study. It's important to adjust for these factors when analyzing the data.

COHORT STUDY DESIGN
Changes in exposure status

- It is also possible for exposure status to change during the course of the study.
- The exposure under study may be subject to variation over time.
- For example, cigarette smokers may quit, or employees may change jobs; therefore, their level of exposure to occupational hazards changes.

Another limitation of cohort studies is changes in exposure over time. To address this, we calculate incidence density. For example, cigarette smokers may quit smoking, or employees in workplaces with occupational cancer risks may change jobs, altering their exposure to risk factors.

COHORT STUDY DESIGN
Analysis

- Collection and analysis of data on the population subgroups, based on exposure, are divided according to variables of interest, like analysis in a cross-sectional study.
- Rates for subgroups are then calculated and compared.
- Data from cohort studies are analyzed in terms of relative risk and attributable risk fractions.

COHORT STUDY DESIGN

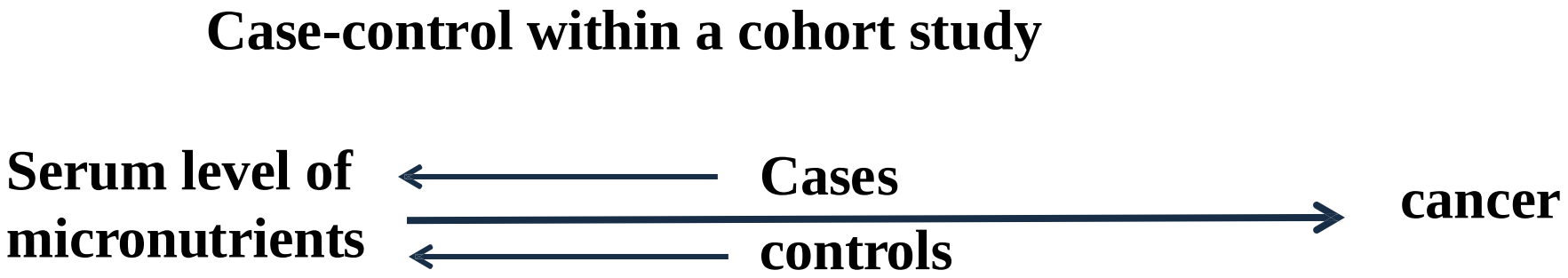
Midpoint analysis

- Occurs when, at a defined point in time in the study, all data collected to that point are analyzed so a decision can be made to stop or continue the study.

For example, in a study examining the impact of working at certain business sites as a risk factor for lung cancer, if we have 10,000 subjects and a 20-year follow-up, we might perform a midpoint analysis after 10 years. If the analysis reveals a significantly higher incidence of lung cancer in the exposed group compared to the non-exposed group, we should stop the study at that point. Continuing the study would be unethical, as it would mean allowing subjects to continue exposure to a known risk factor. The results should be presented, and interventions should be implemented immediately to reduce exposure. Similarly, if we are following people taking aspirin and studying its effect on the incidence of colorectal cancer, and after 10 years we find that aspirin users have a lower incidence of colorectal cancer, we should recommend that the other group also take aspirin. However, if the aspirin group shows a higher incidence of peptic ulcers or upper GI bleeding, we should stop the study and prevent further exposure to aspirin.

If, during a study on the effects of certain pesticides, we discover a high incidence of lung cancer or other cancers in the exposed group after 10 years, the study should be stopped. It would be unethical to continue observing a group at high risk for cancer, and interventions should be made. For instance, if 100 out of 10,000 subjects developed the illness after 10 years, this would be a significant finding, and the study should end, rather than waiting for additional cases to occur.

Nested case-control study



Case-control studies are discussed in the next part of this lecture. Sometimes, while conducting cohort studies—such as on serum levels of micronutrients or smokers versus non-smokers—you may discover cases of a rare disease during the follow-up. In such cases, you can use these rare disease cases to conduct a case-control study (remember, rare diseases are ideal for case-control studies). For example, in a cohort study of pregnant women followed for 20 years, if you identify 20 cases of congenital heart disease, you can conduct a cohort-based case-control study on these cases. This approach is beneficial because you already have all the baseline information and data about the mothers since the start of pregnancy. You will also have detailed information about various risk factors and complete medical records. Essentially, this is a two-in-one study, combining the cohort design with a case-control study to investigate rare diseases.

Framingham Heart Study

Approximately 5100 residents of this Massachusetts community (USA) **are followed for > 30 years.** It was a huge study with a high cost.

Selected because of a number of factors has permitted assessment of the effects of a wide variety of factors on the risk of numerous diseases

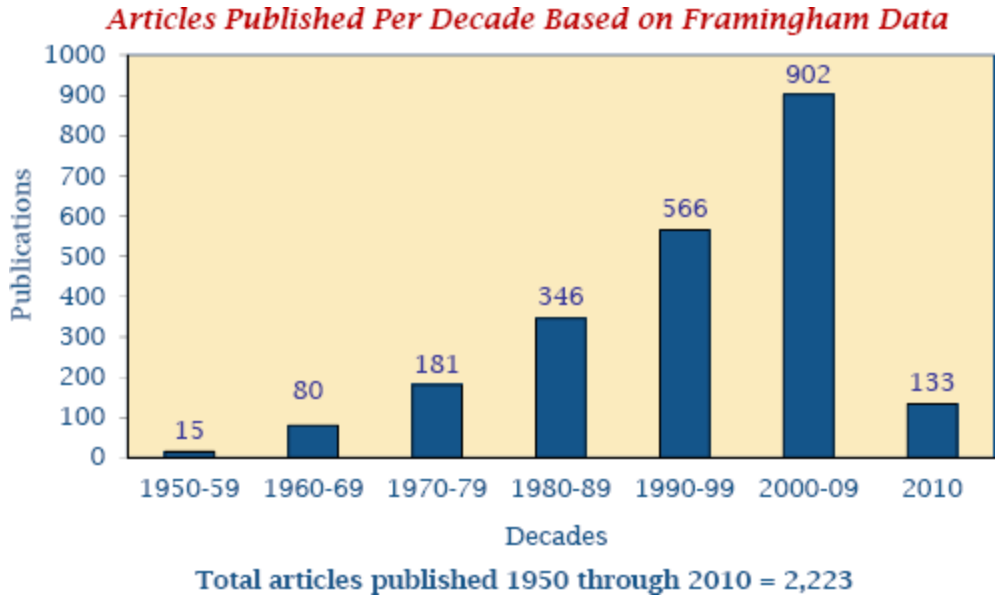
- **stable population,**
- **had a number of occupations and industries represented**
- **had a single, major hospital that was utilized by the vast majority of the population**
- **prepared annually updated population lists that would facilitate follow-up,**

Diseases studied included:

- **coronary heart disease**
- **rheumatic heart disease**
- **congestive heart failure**
- **angina pectoris**
- **intermittent claudication**
- **stroke**
- **gout**
- **gallbladder disease**
- **a number of eye conditions**

The Framingham Heart Study

Notice that this particular cohort study resulted in a large number of publications. Although cohort studies require a significant amount of time, the outcomes they produce are highly valuable. It's important to be patient for the sake of your patients, as the results will be worthwhile in the end.



http://www.framinghamheartstudy.org/risk/index.html

- http://www.ajconline.org/article/S0002-9149(00)00726-8/abstract

COHORT STUDY DESIGN: Summary

- **In general, can investigate the effect of only a limited number of exposure** and birth cohort studies are an exception.
- **Useful for investigating a range of outcomes associated with only one exposure** from a single exposure we can study many diseases.
- **Useful for the study of rare exposure**
- **Not suitable for the study of rare diseases** where we use case-control studies.
- **Follow-up studies are often large and** expensive but they give great outcomes.
- **May take many years to complete** and it's the reason why we sometimes do retrospective or ambidirectional cohort studies.
- **Cannot test current hypotheses**
- **Can measure disease incidence**

Can be used to study the burden of rare illnesses or evaluate their epidemiology because we can't do cross-sectional studies on them to look at the incidence.

Bradford Hill Criteria

1. Strength of the evidence
2. Order in time
3. Consistency
4. Plausibility
5. Specificity
6. Biological gradient
7. Coherence
8. Experiment
9. Analogy

Identifying a risk factor based solely on the relative risk isn't enough to conclude that it causes the illness. Other criteria must also be met. For example, even if the relative risk for smoking and hypothyroidism is significant, we can't immediately say smoking causes hypothyroidism. We need to consider additional factors. This category is crucial when discussing causation, as sometimes a high relative risk is observed, but there may not be a biological justification to support this finding.

فليسَ يَجني ثِمارَ الفوزِ يانِعةً

من جنَّة العلمِ إلَّا صادقُ الهمَمِ

VERSIONS	SLIDE #	BEFORE CORRECTION	AFTER CORRECTION
V1→ V2			
V2→V3			

امسح الرمز و شاركنا بأفكارك لتحسين أدائنا !!