MEDICAL RESEARCH

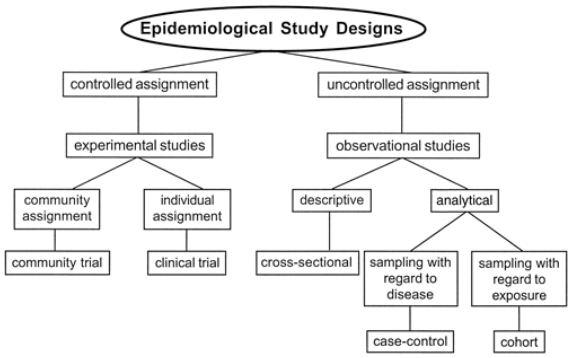
Writer: Tasneem Alremawi

Corrector:

Doctor: Munir Abu-Helalah

Week 6, part 1 (Cohort studies)

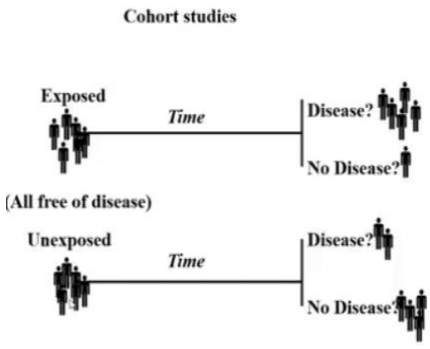
This week covers **analytical studies,** we covered previously observational, descriptive studies such as: cross-sectional studies surveys, ecological studies, case reports and case series, these studies help us to **generate hypothesis** about potential risk factors;



for example, we have seen that a group of patients with cervical cancer, 17 out of 20 in this case series were human papilloma virus positive, and in the cross-sectional study the prevalence of HPV positivity was high among these cervical cancer patients compared with the general population or patients coming with other illnesses. Other observations were that stomach cancer and Helicobacter Pylori could be risk factors for stomach cancer, while smoking could be a risk factor for diabetes, hypertension, hypothyroidism or cancer, we found that there is a high prevalence of these potential risk factors (observations) among patients with different illnesses, and after that **we need to do analytical studies to prove or disprove this hypothesis.**

We have two types of analytical studies, and we described them briefly during the previous weeks, cohort studies and case control studies.

The classical example of analytical studies is the cohort study, we have two groups exposed and unexposed groups at the baseline, both should match in age and gender and not have the disease that we are studying, and then follow them for 10, 20 years to look at the incidents of different illnesses.



For example, if we have two groups, A smokers and B nonsmokers, then we followed them up to look at the incidence of ischemic heart disease, diabetes, cancer different respiratory illnesses etc... and we found that the incidence of ischemic heart disease among the smokers were 100 per 10000, and the incidence for the nonsmokers were 20 per 10000, this means that the relative risk is five (100/10000 / (200/10000)), smokers are at five times higher risk to develop the disease compared with nonsmokers.
This is the general idea of cohort studies, we have exposed group, those with that certain risk factor, and unexposed groups are the non-exposed ones with a match for age and gender, for example if we have a group from general population their age is 20 to 40 and are smokers, the controls should be people living from the same geographical area and their age ranges from 20 to 40, there should not be significant differences at baseline in the age or other risk factors that would lead to variations in the incidence.
We can start with different risk factors: smokers and nonsmokers, HP virus positive HP virus negative, people who are taking aspirin and those who are not taking aspirin for other reasons not as a part of the clinical trial but for other indications such as having for example a risk of heart disease or to prevent recurrent MI or people taking high fiber diet physically active and physically inactive people, we can look at all potential risk factors, different occupation risk factors, lifestyle risk factors etc, so **the good thing about cohort studies is that we can study many illnesses for a single risk factor.**

The key condition for cohort studies is that at the baseline the exposed and unexposed groups both should not have the disease of interest, because we want to compare the relative risk for exposure on the development of that disease.
For example, if we want to see the effect of smoking on the development of type two diabetes, at the baseline we need to do two things, first to ask the exposed and unexposed group if they have diabetes, impaired glucose function or pre-diabetes, if they do they should be out of the study, the second thing we need to is to do tests

and check their glucose profile at baseline to ensure that they have normal glucose function tests and normal fasting glucose HbA1c. Another example, if the study was about the relation between hypothyroidism and smoking, we should ask at the baseline if they had a history of hypothyroidism and screen TSH levels. And if we are looking for the risk factors for breast cancer, at the baseline we should screen patients according to their age group, if they were younger than 40 they will have MRI, and above 40 will have mammogram. So, **at baseline there should be no disease**, and then we follow up.

The good thing about cohort studies is that during the follow up we can calculate the incidence during the study and then we can calculate the relative risk.

How to calculate the incidence?
If we have for example 100 patients who developed hypothyroidism over 10 years (remember the incidence is calculated **annually** as we said in the previous lectures), this means that we had 10 cases per year, so the instance will be 10 per 10,000 per year, and we always prefer to use it as per 100,000, so we will have 100 cases per 100,000 per year, and if it was for six months, the incidence should be multiplied by two so that we'll have the annual incidence, for example, if we detected 50 cases in 6 months, there will be 100 cases annually, and if the follow up was for two years and the incident was 100 cases, it will be 50 per the number of population per year.

The great thing about cohort studies is that we can establish **temporal relationships,** meaning that we are confident that the exposure or the risk factor was before the disease, since we did ask them and did tests at the baseline to ensure that they didn't have the disease.

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)

Cohort study is an uncontrolled assignment, where investigators don't do anything for the subjects they only observe and compare groups on the long term.
For example, if I want to start to assist the impact of daily intake of multivitamins on different diseases, if I give the subjects these multivitamins this is going to be a clinical trial not a cohort study, but if I'm just comparing subjects who are taking these vitamins routinely already with another group who don't take them then this is a court study.
If I want to look at the impact of taking aspirin on the incidence of colorectal cancer as a preventive factor, and I go to the hospital records and I find subjects who are taking aspirin regularly and I follow them and compare them with controls from the Jordan Universe Hospital who are not taken Aspirin regularly then this is a cohort study, but if I invite the subjects to receive aspirin or to receive Placebo, then this is a clinical trial.
In court studies, case control studies similar to the other descriptive studies as an investigator I should not intervene with anything dose or frequency, so these groups will be followed over time to determine whether the incidence of a particular disease in any group is higher or lower in the exposed group or the non-exposed groups. The incidence will be higher in potential risk factors such as smoking, and lower in taking preventive factors such as aspirin.

Cohort study: examples: (there are many other than just those)

- **Life expectancy of cerebral palsy children.**
- **Fine needle breast biopsy and breast cancer.**
- **Aspirin intake and colorectal cancer.**
- **Hypertension as a risk factor for spontaneous intracerebral Hemorrhage.**

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

- **Rare disease: we conduct case control study starting with cases.** (discussed in the 2nd part of this lecture)
- **Rare risk factor: we conduct a cohort study starting with rare risk factors.**

In rare diseases we will use case control studies and **in rare risk factors we use cohort studies.**

Because otherwise it is going to be hard to find it, for example if we need to study the impact of radiotherapy for children who were diagnosed with childhood malignancies on the development for adult malignancies or the impact of fine needle aspiration for benign malignancies on the development of breast cancer later on, it's going to be difficult to find these risk factors in case control studies, if you interview patients with certain illnesses, you will not find these rare risk factors, so if you have a rare risk factor and you want to study the impact of it on different illnesses you need to start with that risk factor, for example you have children with malignancies and radiotherapy you go to the records of the hospitals where the patients with childhood malignancies are treated and you look at the data there and you make the follow up for the children and compare them with children without radiation therapy or without childhood malignancy and if you want to make a study about the effect of fine needle aspiration on the development of breast cancer, you'll search through the histopathology records and look for women who had fine needle aspiration and their results were benign and compare them with women coming to the same hospital but without having needle biopsy and compare the incidence of breast cancer among these two groups.

While in rare diseases, it's best to do case control studies, because since it's rare the incidence of it will be very low, so you will need a huge sample size to find a good number of cases and to identify the cases as well, for example if an illness had the incidence of one per 100,000, and if you want to do a cohort study you will need 100,000 subjects to be followed for 10 years to get 10 cases of that illness, so it's not feasible nor efficient.
The general idea is that if we want to study a rare disease or risk factor we will start with subjects who have it from the baseline, for rare risk factors we use cohort studies to start with subjects who already have that risk factor, and for a rare disease we use case control studies so that the subjects would already have that rare disease from the start of the study.

Remember we said that in cross-sectional studies, if you have a rare disease and you want to study the magnitude of that rare disease, we can't use cross-sectional studies because it would need a huge number of participants to get enough number of cases.
To study the burden or magnitude of a rare disease or a disease with short duration, we conduct court studies, for example if you want to look at the magnitude or burden of congenital heart diseases in Jordan, you should use a cohort study to calculate the incidence not the prevalence, through the follow up of pregnant women in Jordan over the next two years and look at the incidence of how many numbers of cases of congenital heart diseases happened to their babies, but if I want to study risk factors for congenital heart diseases it is best to use case control study.
In any medical textbook, in the introduction, epidemiology, of common disease it starts talking about the prevalence of that disease, in rare diseases it talks about the incidence of that disease, so we always assess **the burden of rare illnesses by calculating the incidence using the cohort studies.**

Cohort study: Primary purposes

- **Descriptive (measures of frequency)**

– To describe the incidence rates of an outcome overtime, 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** (for example to calculate the incidence of gastroenteritis, influenza infections or MI over 1 year time we need to use cohort study, because it would be difficult to use cross-sectional studies for a short period 6-5 months, so cross-sectional studies can't study seasonal variations)
- **Objective measures of exposure, such as biological markers, are preferred over subjective measures.**

Cohort studies are used to assess biological markers, we look at smoking, pre-diabetes and type two diabetes (biological), they are preferred over subjective measures (**because subjective measures can vary over time while biological markers won't**), for example if you have a question to identify subjects with depression (subjective) and you make a follow up for them this can be hard because this is something that can vary over time that's why we prefer to use biological markers rather than subjective measures in cohort studies.

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,** remember to make sure that the subjects don't have the disease of interest at the baseline.
- **Not open to bias as much as other types of study,** in case control studies we have interviewer, selection, recall bias etc. which are discussed in the next part of this lec.
- **Direct calculation of the risk ratio or relative risk is possible,** in case control studies we can't calculate relative risk because we only have the number of cases who have that disease and we don't have the number of the general population we don't have the dominator to calculate the risk factor and we can't calculate the incidence too, so we calculate the odds ratio not the relative risk.
  Relative risk is the ratio of the risks for an event for the exposure group to the risks for the non-exposure group.
- **Provide information on multiple exposures.**

Cohort study design: Limitations

- **Not efficient for rare diseases.** If the incidence was lower than one per 10,000 this is rare disease, and we can't calculate it because we need a huge sample size.
- **Can be expensive and time-consuming.**
- **Large sample.**
- **Drop-out biases**, because of the large sample size, some might change their location or get lost during the follow ups and the risk factors can change over time (smokers may give up smoking and nonsmokers start to smoke) which can actually be adjusted by the incidence density, but there will still be limitations for the follow up studies.
  If study goes over many years, can get considerable loss to follow up. This can 'dilute' results or lead to bias, and therefore the validity of result 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.**

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.

Physical activity	Cognitive impairment		
	Yes	No	Total
Moderate (Non exposed)	10	990	1000
None (Exposed)	100	900	1000
Total	110	1880	2000

This is an example of a cohort study showing the impact of physical activity and incidence cognitive impairment, there are 3903 participants older than 55 enrolled during that period for follow up for two years and physical activity was classified as no activity, moderate and high activity and cognitive impairment was assessed to assess the impact of physical activity on cognitive impairment.
People with no physical activity represent the exposed group (Lack of activity is the risk factor) and people with moderate physical activity don't have 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 to have cognitive impairment compared with physically active subjects.

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

It is proportion (0 - 1)

The incidence is calculated over a period of time, so if we have 100 cases over two years, there will be 50 per year, and if we have 50 per six months, it will be 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, if you have broken stairs (we call them a hazards), you might get injured more than if they were not broken.

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 <u>no association</u> between exposure and disease.

- **RR>1**

  Exposure is associated with an <u>increase</u> of the frequency of the disease.

- **RR<1**

  Exposure is associated with a <u>decrease</u> of the frequency of the Disease. (called a <u>preventive factor</u>)

•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.

	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)

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

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.

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

This is another example about the rate of risk, you can see that women taking contraceptives are 1.4 times at higher risk to have bacteriuria, and something very important is to look at the **confidence interval.** Remember that in t-tests we look at the zero as the starting point and if the confidence interval was minus one to three, it will pass through zero and it's not going to be significant. But here in the relative risk or other ratios, **the relative risk reference point is 1 (it should not pass it), if something is higher than one this means that it's a relative risk and if it was less it won't be significant,** so if the confidence interval was 1.2 to 1.8 then this is considered as a significant factor, but if it was 0.9 to 1.6 it is not going to be significant.

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 $I_{exposed} < I_{unexposed}$

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

If RR is higher than one this mean that it's a relative risk factor, and if the RR was less than one then it's a preventive factor.
Preventive fraction is the incidence of unexposed minus the incidence in the exposed group over the incidence in the unexposed.

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

	IHD risk
Exercise	2/100
No exercise	8/100

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

Preventive fraction is 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 the relative risk here, considering the no exercises as the risk factor (exposed group), it will be 8% over 2% = 4, meaning that the lack of physical activity causes 4 times higher risk to develop IHD in comparison with subjects who are physically active.

If you want to assess the impact of exercise to prevent ischemic heart disease, we use the preventive fraction, so we have the incidence of exposed 8% minus 2% over 8% which gives us.75, as a percentage it's 75% of cases of ischemic heart disease don't exercise and can be prevented by exercise.
But we can't say that we can eliminate the whole risk and that the 8% of not exercising all are responsible for the ischemic heart diseases because we have other risk factors, so the difference here (the risk of developing IHD from not exercising) is 6% not 8%, it is true that not exercising is an important risk factor for ischemic heart disease incidence among this group, but it is not the only one and we have other risk factors that contribute to the 8% that we see in the table.
Also, you can see that with exercise the incidence was not zero, it's 2% so we don't compare the group who didn't exercise with zero we compare it with two and that is why we said that the percentage is 6% not 8% (percentage of developing IHD from not exercising and other risk factors – percentage of developing IHD from other than not exercising = percentage of developing IHD from not exercising (8% - 2% = 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.**

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.**

• 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 to have **cooperative subjects** who are staying permanently, not someone who comes to work in a certain area for a period of time and go and include them in a 10 year follow up.

Also, we should have what we call **inclusion criteria** which determines who should be in the study based on gender and for example smokers who smoke at least once a day, so if you have someone who smoke every other day, they will not be part of the study and ex-smokers also will not be in the study.

The degree of surveillance should be similar between the exposed and the unexposed group, this means that they both should fit in the inclusion criteria and then become identified as smokers and nonsmokers, and you need to have the new criteria that the control should not be ex-smokers because they might also be at risk of some illnesses and you should have the same frequency of examination, every six months for 10 years for example, we need to insure that both at baseline are comparative groups with no significant differences in age, gender and other risk factors as well, what we call **confining factors.**

Types of cohorts:

- **Birth cohort: all individuals in a certain geographic area born in the same period (usually a year),** we follow up subjects who have the same birth year and place, here we can study many

risk factors for many illnesses, not just a single risk factor like exposure cohort, the key thing about birth cohort is that at the baseline there should be a full list of all the information about the subjects, the risk factors and illnesses that are going to be studied, e.g. a study for physical activity, fat intake (dietary factors) and smoking (social demographic factors) etc. for people born in the year 2000 or a study for people aged 20 who live in the same city and follow the up 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 follow up workers at the same facility or med students from the same uni to look at 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 work place ,for example in JU or JUH, where after a follow up for 20 years people with severe illnesses or disabilities would have changed their jobs, lost it or left it, so now the sample won't be representative anymore for the whole population, that's why we should always have **samples from the general population not from a healthy workers place,** previously in the 70s 80s there were several cohort studies coming from work places, but we don't encourage this anymore.

Cohort study design

- **Measurement of exposures should be based on intensity, duration, regularity, and variability,** for example in a study for

physical exercise or smoking, we need to know the duration, frequency of it etc.

- **Some exposures are acute, one-time episodes never repeated in a subject's lifetime,** as in people who received certain medications, e.g. children who received radiation therapy due to childhood malignancies and we followup even though they had the same exposure with no change over time.
- **Other exposures are long term, such as cigarette smoking or use of oral contraceptives** or exercise.
- **Exposures may also be intermittent.**

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,** and there might be some missing or incomplete data at the base line and throughout the duration of the study, so retrospective studies have a limitation that the data could be incomplete, in this case it is better to do it prospectively as a classical cohort.
- **The follow up was completed in the past, therefore, we call it a retrospective cohort study.**

As we said before the classical type of cohort studies is when we have exposure and non-exposure groups and we follow them up, e.g. in a study about **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 note for non-smokers,** if we already have data from previous studies, it shouldn't be done as prospective,

because it will be time consuming and we can do the study quickly as a retrospective study, but how can we do that?
The family medicine department would have the record files or we use the general practice records and look for the records for that certain duration and then split the files at the baseline into two groups, medical notes of smokers and medical notes of nonsmokers, **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.**

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 additional 12 year. This means a combination of retrospective and prospective data.**

So, imagine that we are in 2015 and we wanted to do a retrospective study for a period of 20 years, and we started looking at the files but we noticed a problem that from the year 2002 and before the data was not enough for the study and we can't just use the data that we found between 2015 and 2002 because we want that 20 years of duration, so the solution here would be to use the data that we have retrospectively and we start collecting data prospectively too, so that by the year 2023 we would have another 8 years of duration and the data collection would be complete, collected both retrospectively and prospectively and this type is called ambidirectional cohort study. These studies are good especially in developing countries because we don't have the complete records from the past, so if you didn't have enough data from a certain year to do a study

retrospectively you would start with the complete records that you have from the previous years and then you make a follow up for the future so that you can save years from the retrospective study that you didn't have enough data in.

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 for cohort studies is being unable to study rare diseases because you need a huge sample size for a long duration to get enough number of subjects.

Another issue with cohort studies is the loss during flow-up of subject, they might be not interested anymore in the study, or they changed their location, died from other causes, we need to adjust for all these factors.

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 for cohort studies is changes in exposure and for that we calculate the **incidence density**, for example cigarette smokers may quit smoking and if a work has occupational cancers

employees may change their job and that would change their exposure to the 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 studying the impact of occupation in working at certain business sites as a risk factor for lung cancer, if I have 10,000 subjects and the follow up was for 20 years, at 10 years a midpoint analysis is made and if we find that this risk factor is a significant risk factor with a significant higher incidence in the exposed group compared with the non-exposed group, then we need to stop the study there because we should not have the subjects being exposed to this factor further anymore, and we need to show the results and show that this is the risk factor and we need to do interventions there (as this is the whole goal from the start).

Also, if we are making a follow up for people taking aspirin and the incidence of colorectal cancer, and the duration of follow was 20 years, after 10 years at the midpoint check if you find that people taking aspirin have lower incidence of colorectal cancer, then you should recommend the other group to receive aspirin.

And if you find that the group taking aspirin have higher incidence of peptic ulcers or upper GI bleeding, you will have to stop the study there and not give the exposed group more aspirin.

Or if you are studying the effect of certain types of pesticides and you saw at 10 years that we have a high incidence of lung cancer or that

certain occupations have a high incidence of different cancers, or any other risk factor and you find a significant difference in the midpoint analysis it would be unethical for us to continue the follow up and we should stop it because we know that these subjects are at high risk of that illness, so we would need to stop the study and make interventions, for example if we have 10,000 subjects and 100 of them developed the illness after 10 years, this is considered significant and the duration of follow up should be stopped and we should not wait for an additional 100 patients to develop that disease while we are watching.

Nested case-control study: Case-control within a cohort study

Case control studies are discussed in the next part of this lecture. Sometimes when we are doing our cohort studies, for example on serum level of micronutrients or smokers and non-smokers, during the followup of subjects in your cohort study you discovered cases of a rare disease, then we can use them to also do a case control study on them (remember rare disease= case control), for example in a cohort study for pregnant women for 20 years, after birth you had 20 cases of congenital heart disease in your study then you can do a cohort case control study on them, and they're great actually because you would have all the baseline information and all the information about their mothers since the start of pregnancy, and you would know about the different risk factors and you would have the complete medical notes for them, so here we **conduct a case control study on subjects taking part in a cohort study who have rare diseases**. (2 in 1).

Framingham Heart Study:

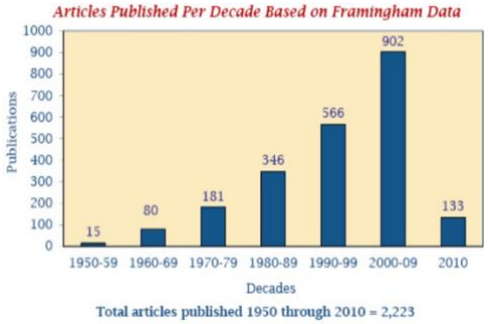
Approximately 5100 residents of this Massachusetts community (USA) **are followed for > 30 years.** It was a huge study with a high cost.

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**



Notice that this one cohort study had a huge number of publications, although cohort studies take a lot of time, the outcomes are going to be great.
(u need to be patient for ur patients صعع)
The website of the study: https://www.framinghamheartstudy.org
The article: https://www.ajconline.org/article/S0002-9149(00)00726-8/fulltext

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 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:

Identifying a risk factor from the relative risk isn't enough to say that it causes the illness as there are other factors that should also be met, can't just say that for example smoking is a cause for hypothyroidism because the relative risk was significant, we should also see these criteria.

This category should be considered when we are talking about **causation**, because sometimes we have a high relative risk, but we don't have the biological justification for this finding.

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

E-learning question: To study risk factors of rare disease, we need to conduct:
Case control study ✓