MEDICAL RESEARCH

Writer:

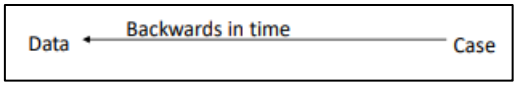
Corrector:

Doctor:

Case control studies

- As mentioned in the previous lecture, the classical way of conducting an analytical study is a cohort study, but since one of the key limitations of them is that we can't use them for rare diseases. This is why case control studies are the appropriate ones when studying a rare disease.
- **They are studies in which a group of people with a particular disease (the cases) are compared with a group of people without the disease (the controls). The purpose of the comparison is to determine whether, in the past, the cases have been exposed more (or less) often to a specific factor than the controls.**
- So, we start with cases and we have match controls patients. For example, we can consider newborns with a certain congenital heart disease as cases and healthy newborns who were born during the same period and at the same hospital as controls.
- We'll start with the disease in case control studies course, however, in cohort studies we start to the risk factor.
- The purpose of the comparison is to determine whether in the past these cases had been exposed to more or -less often- to a specific factor than the controls.
- The classical way to study several risk factors, especially for rare disease, are the case control studies.
- AS a reminder: If you're considering a rare exposure, you need to do a cohort study.
- As an example, if you have 200 patients with lung cancer and we have 600 controls that match in age and gender and you are looking at the childhood exposure to radiation therapy, maybe none of your lung cancer patients had childhood malignancy; so if you have a rare factor then you should better start with a cohort study.
- **This type of study is done to identify factors that could be responsible for the development of a disease or drug use problem.**
- **The direction of time:**
  - **Cases identified now.**
  - **Data on past events collected.**

- Accordingly, in case control studies, you can assess factors that no longer exist.
- They are retrospective studies.



- **Designed to assess association between disease occurrence and exposures (e.g., causative agents, risk factors) suspected of causing or preventing the disease.**
- **To sum and clear up what was mentioned before:**
  - **A group of people with a disease are compared to a group without the disease from the same population.**
  - **Compare exposure to risk factors in both groups.**
  - **Able to look at many different possible risk factors.**
  - **Able to study diseases with a long latency period.**
    - sometimes we have certain environmental factors that people get exposed to at their late teen years or early 20s and they will have the disease in their 50s or 60s; consider such diseases as ones with long latency period and this is applied actually for most occupational cancers so when we conduct a case control study we can have complete environmental occupational history, for instance we can ask patients with lung cancer about their incubation since the age of 18 and then we can study the odds ratio and evaluate these risk factors.
  - **Most common analytic study design seen in the medical literature today.**
  - **In general, the cases included in a case-control study include people with one specific disease only.**
  - **But, a case-control study can provide information on a wide range of possible exposures that could be associated with that particular disease**
  - **Useful for the study of rare diseases.**
  - **Not suitable for the study of rare exposure**
  - **Relatively small and inexpensive.**
  - **Takes a relatively short time to complete.**

- **Can test current hypotheses.**
- **Cannot measure disease incidence.**
  - You can have 20 cases of congenital heart disease to study, but you still don't how many other cases there are in other hospitals.
- **Cases have the disease of interest. (Eg. Cerebral palsy)**
- **Controls do not have the disease. (Eg. Healthy babies born at the same time).**
- **Controls are chosen from the same population yielding the cases.**
- **One key feature of a case-control study, which distinguishes it from a cohort study, is the selection of subjects based on disease status.**
- **More efficient than a cohort study because a smaller sample size is required.**

➢ **Challenges in case control studies:**
- **Selecting cases**
  - **Eligibility**
- **Selecting controls**
  - **Representativeness** (They should match the characteristics of the patient- only difference is that they don't have the disease).
- **Exposure assessment**
  - **Accurate** (If I'm looking at the effect of high aspirin intake during the first trimester of pregnancy as a risk factor for congenital heart disease then I should have accurate information about the dose of aspirin taken in the first trimester).

➢ **Design of case control studies:**
- **Comparability: Two groups must be as similar to each other as possible so selection of controls is very important. Controls must be as similar as possible to cases – except that they do not have the outcome (disease).**
- **Outcome (disease) must be very clearly defined- Case definition. (Diagnostic criteria must be clear).**

- **Use objective data about exposure status wherever possible, to reduce the risk of bias.** (You need to go back to medical records to collect precise information about the aspirin dose the mother was taking in her first trimester 'Don't depend on the patients' memory').

➢ **Strengths:**

- **Suited to study disease with long latency periods, but can be used in outbreaks investigations.**
- **Optimal for rare diseases.**
- **Efficient in terms of time and costs: relatively quick and inexpensive.**
- **Allows for evaluation of a wide range of possible causative factors that might relate to the disease being studied.**
- **Odds ratio estimated.**

➢ **Limitations:**

- **Very susceptible to bias (especially selection** -how you select your cases, controls and your case definition- **and recall bias**- sometimes patients don't remember certain information accurately-**) as both the disease and the exposure have already occurred when participants enter the study. Also, cases and controls might not be representative of the whole population.**
  - We also have what's called interviewer bias; simply explained, imagine someone making interviews on the case of congenital heart disease, he had two groups-cases and controls- when interviewing the mothers of babies with congenital heart disease he tried to make a detailed interview with them, however, when he interviewed the mothers of healthy babies, he was just asking the question quickly, skipping and not going through the records care because he knew they are healthy.
  - how we can avoid interviewer bias? By blinding the interviewer as he shouldn't know which group is which.
- **We cannot calculate incidence or prevalence rate of disease.**
- **We cannot be certain that exposure came before disease**
- **Choosing controls is difficult.**

- **Controls do not usually represent non-exposed population** (Sometimes we prefer to have controls from the same hospital and others from the general population who match for age and gender).
- **Past records incomplete.**
- **No absolute risk estimates (We can't calculate the incidence and relative risk, only odds ratio).**

- **Data analysis: Data collection and analysis are based on whether the case-control study involves a matched or unmatched design- The measure used typically in case-control studies is the odds ratio.**

Odds Ratio

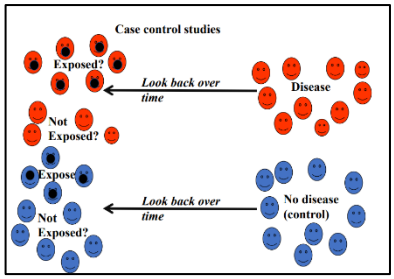
- **odds of a particular exposure among people with a specific condition divided by the corresponding odds of exposure among people without the condition under study.**
- **The word "odds" means the chances of an event to happen. The Odds of an event is the ratio of the event to happen over the event not to happen.**

$$Odds(A) = \frac{probability(A\ happens)}{probability(A\ does\ not\ happen)} = \frac{prob(A)}{1 - prob(A)}$$

$$prob(A) = \frac{Odds(A)}{1 + Odds(A)}$$

$$OR = \frac{\text{Odds of exposure}_{\text{cases}}}{\text{Odds of exposure}_{\text{controls}}}$$

	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



- **Odds of being ill in exposed=a/b**
- **Odds of being ill in non-exposed =c/d**
- **Odds ratio (OR)=Odds in exposed/Odds in non-exposed = OR=(a/b)/(c/d) =ad/bc.**

Data from a case-control study of current oral contraceptive (OC) use and myocardial infarction in premenopausal female nurses

	Myocardial infarction		
	Yes	No	Total
Current OC use			
Yes	23	304	327
No	133	2816	2949
Total	156	3120	3276

Data from L. Rosenberg et al., Oral contraceptive use in relation to non-fatal myocardial infarction. *Am. J. Epidemiol.* 111:59, 1980.

$$OR = \frac{23 \times 2816}{304 \times 133} = 1.6$$

- **Women who were current OC users had a risk of MI 1.6 times that of non-users.**
- The following table shows the numbers of participants in a case-control study assessing early life exposure to diagnostic radiation and ultrasound scans and risk of childhood cancer:

Radiation	**Case**	**Control**	Total
Yes	140	165	305
No	1550	5693	7243
Total	1690	5858	7548

- **Odds of outcome in exposed = 140 / 165 = 0.85, Odds of outcome in non-exposed = 1550 / 5693 = 0.27**
- **Outcome odds ratio = (a/b) / (c/d) = 0.85/0.27=3.1**
- This means that children exposed to radiation have 3 times the risk of cancer compared to the ones who were not exposed.
- **Methods of data collection:**
  - **Case-note review: Completeness.**
  - **Postal questionnaire: response rate.**
  - **Interview: Detailed information.**

- **How many controls?**
  - **control-to-case ratio is 1:1**
    - **this is the optimal when the number of available cases and controls is large and the cost of obtaining information from both groups is comparable**
  - **control-to-case ratio is 1: n**
    - **When the number of cases is limited or when the cost of obtaining information is greater for cases or controls.**
  - **As the number of controls per case increases, the power of the study also increases.**
  - **It is not recommended that the ratio increases beyond 4:1** (This is the ideal situation actually because controls usually don't have the risk factor and we want to have a fair comparison).

- **Selecting Cases and Controls:**
  - **Identification and collection of cases involves specifying the criteria for defining a person as a case—in other words, as having the disease (also called case definition).**
    - **This definition consists of a set of criteria, also called eligibility criteria, for inclusion in the study. There are also criteria for exclusion from the study.**
  - **The next step is selection of the controls.**
  - **Controls are chosen from the source population.**
  - **The source population is usually defined by geographic area. It is important to select controls so that participation does not depend on exposure.**

Source of controls:

- **The ideal situation is a random sample from the same source population as the cases.**
- **Investigators may use more than one control group.**
- **Controls can be selected by sampling: The general population in the same community; the hospital community (patients in the same hospital); individuals who reside in the same block or neighbourhood; and spouses, siblings, or associates (schoolmates, co-workers) of the cases.**

- **Examples on obtaining cases and controls for case-control studies:**

Study	Source of cases	Source of controls
PROM (premature rupture of membrane)	Hospital patients	Hospital patients
Rheumatoid arthritis	Outpatient clinic	Other outpatient clinic
Cervical screening	GP register	GP register

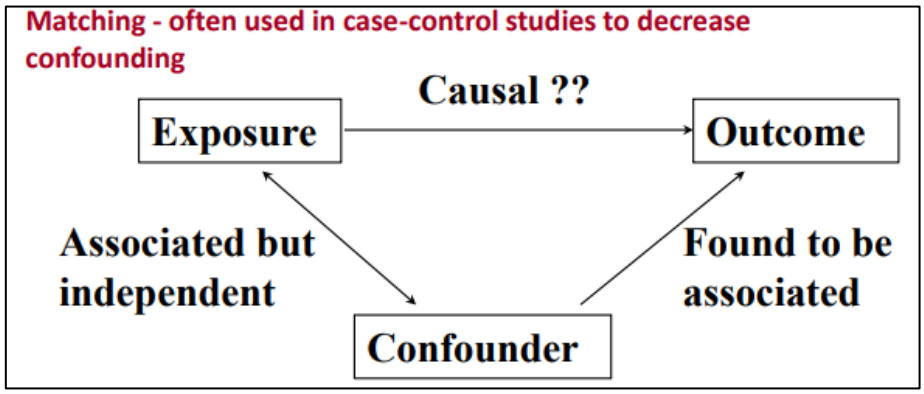
- **Matching Cases and Controls:**
  - Matching is a popular approach to control for confounding and selection bias in case-control studies.
  - Matching cases and controls helps to ensure that these groups are similar with respect to important risk factors, thereby making case-control comparisons less subject to confounding or selection bias.
  - For example, if age and sex are the matching variables, then a 35-year-old male case is matched to a 35-year-old male control
  - Pair matching (one to one individual matching)
  - The use of matching usually requires special analysis techniques (e.g. matched pair analyses and conditional logistic regression).
- Once cases and controls are selected, information must be collected on prior exposure to the risk factor(s) of interest.
- Interviews and questionnaires are the most common means of determining a subject's exposure history and medical records review is another source.
- The most objective means for characterizing exposure is the use of a biological marker.

Bias

- **Bias is any systematic error in an epidemiological study that results in an incorrect estimation of the association between exposure and risk of the outcome.**
- **Selection bias: inappropriate controls.**
- **Observation bias:**
  - **Subject and recall bias: eg recall bias of mothers with cerebral palsy babies.**
  - **Interviewer bias: blind if possible.**
  - **Misclassification** (You should consider case definition and eligibility criteria to avoid misclassification).

Confounding

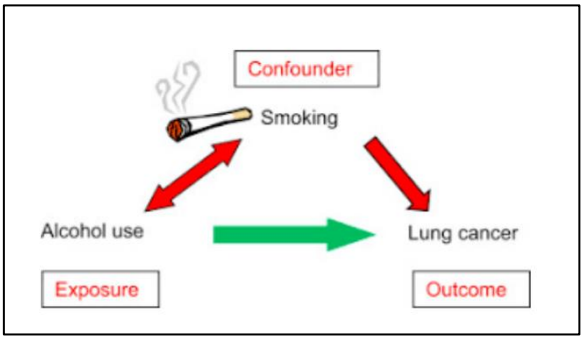
- **A confounding factor is one that is associated with the exposure and that independently affects the risk of developing the outcome, but that is not an intermediate link in the causal chain between the exposure and the outcome under study.**



- Let's say that a case-control study showed that the odds ratio for lung cancer and heavy alcohol intake was 10, this means that heavy alcohol drinkers are 10 times more prone to develop lung cancer compared to the ones who don't drink. So, alcohol here is a risk factor, but in reality, it's a confounding Factor.
- If you take smoking into consideration, you can decide whether alcohol or smoking is the actual risk factor. We'll go to these heavy

alcohol drinkers and we split them into two groups: heavy alcohol drinkers+smokers and heavy alcohol drinkers+nonsmokers.

- You'll find that in heavy alcohol drinkers+nonsmokers the OR drops down to one-Hypothetically- which is not significant, while in the group who are heavy alcohol drinkers+smokers the OR increases actually so it's smoking that's responsible for lung cancer.



- P.S. : Skip the figure above -This in not the real situation-; the doctor just wants you to understand the concept and know that if we evaluated the association between alcohol and lung cancer then found out that smoking also has a role then we might think that smoking is the confounder.