Epidemiological study methods

2) Case-Control studies

Design

A case-control study is an important study design which is relatively quick and at the same time useful to search for causes of a disease or a health-related problem. The basic element of a case-control study is comparison of cases of the disease with people without the disease (controls) in relation to the risk factors of interest. To do this, we first select a certain number of cases and then select a similar number of controls and compare their risk factors. Therefore, the starting point in a case-control study is the outcome (disease).

In a case-control study we don't have information about the total population at risk; we only have information on a certain number of cases which we have selected. Therefore we can not calculate risk of disease in the population. Instead, we calculate odds. We calculate odds of exposure among the cases and odds of exposure among the controls and we compare these two odds to obtain odds ratio.

For example a case-control could be done to investigate the association between smoking and lung cancer. We select a number of individuals who have (had) lung cancer and select a similar number (usually but not always) of comparable individuals who have no lung cancer. We then collect data on smoking (the main risk factor) and other potential risk factors among the two groups and compare odds of smoking and other risk factors between the two groups.

Why we conduct case-control studies?

Case-control studies could be used for the following purposes:

1. To study etiology of diseases and health-related problems: by comparing cases of a disease with people without the disease, in relation to potential risk factors, we can see which risk factors are more likely to be associated with the disease. Analytical methods can investigate the effect of several risk factors separately and provide a better understanding for the causative role of each risk factor
2. To study etiology of rare diseases and diseases with long incubation periods: diseases that have long incubation periods are difficult to investigate with other research methodologies because we would need a long time to wait. Since a case-control study investigates a disease after it has happened, it is very useful for investigation of diseases that take a long time to develop such as cancers.
3. To study vaccine effectiveness: another use of case-control studies is investigation of vaccine effectiveness. By comparing people who are vaccinated and those who are not vaccinated against a disease, we can estimate the effectiveness of a vaccine against the disease it is designed for.

Steps in undertaking a case-control study

1. Defining the study question

Since case-control studies are usually used to investigate etiology, the study question is usually about identification of risk factors. For example we may ask "is smoking associated with lip cancer?" "Is poor parental education a risk factor for childhood injuries?" Or we may try to identify potential risk factors of an unknown condition such as "what are the risk factors for

death in patients admitted for burn injuries?" In vaccine studies, the question is about effectiveness of the vaccine; therefore we may ask "is BCG vaccination effective in protecting people from TB"?

2. Selection of cases

When we have agreed on the study question, we go on to select the cases. From the sample size calculation, we know how many cases we need and how many controls we need. Before starting the recruitment of the cases, we have to take into consideration the following issues:

- **Case definition:** we must have clear criteria to define which individuals within the population will be included as cases. All cases must be selected according to these standard criteria. Usually we develop inclusion criteria and exclusion criteria, inclusion criteria tell us which people could be cases and exclusion criteria tell us which people should not be recruited as cases. These criteria are important and should be followed strictly on both cases and controls.
- **Source of cases:** we must define the population from which the cases are drawn, because the controls will need to be selected from the same population. For example if cases are selected from residents of Sulaimani city, controls should also be residents of Sulaimani city.
- **Incident or prevalent cases:** Recruitment of cases can be retrospective (prevalent cases) or prospective (incident cases). Retrospective recruitment selects the cases from the past i.e. cases which have already happened, whether alive or dead. For example in a study of melanoma, we could select cases of melanoma who have been diagnosed, treated, or even died. Most case-control studies use retrospective case recruitment. Prospective case recruitment means selecting cases prospectively when they develop during a certain period of time. For example in a case-control study on risk factors of childhood burns, cases were selected from new burn injuries attending the burns center during the data collection period.

3. Selection of Controls

Selection of controls could be one of the most difficult stages of a case-control study. This is because if controls are not selected in a proper way, the comparison with cases will not be appropriate and the results will be biased. Controls should be selected from the same population which has produced the cases. The same exclusion/inclusion criteria should be applied to controls; in short controls will be similar to cases but they will not have the outcome in question. The following issues must be kept in mind while selecting controls:

- **The source of controls:** In theory, it is always better to select controls from the population (population-based controls) because they will be more representative to the population. However, selecting controls from the population is more difficult and costly. In practice most case-control studies select controls from hospitals (hospital-based controls). This is because cases are usually selected from hospitals and the use of hospital-based controls is more convenient for researchers. When the cases are derived from a health facility, proper controls from the same health facility would be easier to recruit. However, these controls may not be representative of the population which produced the cases with respect to one or more exposures of interest causing selection bias.

- **Number of controls per case:** we usually select one control for each case but we can select **more** than one control for each case, for example 2 controls for each case, or even three. We can increase the number of controls to compensate for shortage of cases.

- **Matching:** controls may be matched to the case. Matching is selecting a control to each case which is similar with respect to one or more exposure. For example if we match by sex, this means when we find a male case, we have to select a male control (one to one matching). So that at the end of the recruitment we will have the same number of males and females. The most commonly matched for variables are age and sex. The objective of matching is to control for the effect of that variable on the outcome. Frequency matching is also possible in which we don't match one to one, but we do overall matching by taking similar percentages of the variable (e.g. males and females) in both cases and controls.

4. Data collection

The methods used to measure exposure will depend on the type of exposure, and the period of time when the potential exposure may have occurred. This can be through:

- Personal interviews: face-to-face interview with individuals using a questionnaire
- Postal or telephone interviews using a questionnaire
- Medical records: previous medical records of the individuals
- Physical examinations
- Diagnostic tests

5. Data Analysis

In case-control studies we calculate odds and odds ratios for the exposures of interest. We calculate odds of exposure in cases and odds of exposure in controls. Then we compare the odds of exposure in cases with the odds of exposure in controls (odds ratio). The classical way to display data from a case-control study is shown below.

Exposure	Outcome	
	Case	Control
Exposed	A	B
Un-exposed	C	D

Odds of exposure among cases= cases exposed / cases not exposed= A/C
Odds of exposure among controls= controls exposed / controls not exposed= B/D
Odds ratio= odds of exposure among cases/ odds of exposure among controls
= A/C / B/D =AD/BC

Example: in a case-control study to investigate association between poverty and childhood burns in Sulaimani, 248 cases and 248 controls were studied. Fifty four cases and 16 controls were from poor families. Calculate the effect of poverty on childhood burns.

	Case (Burn injury)	Control (No burn injury)
Poor	54	16
Not poor	192	232

Odds of exposure among cases= a/c= 54/192=0.28
Odds of exposure among controls= b/d= 16/232= 0.07
Odds ratio= 0.28/0.07= 4.0
We can also calculate directly using
Odds ratio= ad/bc= 54*232/16*192= 12528/3072 = 4.08

What does this mean? This means that the odds of being from a poor family was 4 times more in children with burn injuries compared to children with no burn injury. This is an indication of association between poverty and burns.

6. Interpretation of results

Before interpretation of results we must remember that there are possible sources of error which we should be aware of. Both selection bias and information bias can happen in case-controls studies. Gross bias may cause the results of the study to be useless. In addition we have to remember strengths and limitations of case-controls studies.

- Selection bias: selection bias is related to the way we have selected the cases and the controls. If the cases and controls are not similar or if the controls are not representative to the population that produced the cases, selection bias is likely.

- Information bias: Information bias is related to the information collected; any error in the measurement of exposure or outcome that results in inaccuracy of information collected between cases and controls is called information bias. One type of information bias is *reporting bias* when the individual providing the information makes mistakes either because he/she does not remember well (recall bias) or for any other reason. The other type of information bias is called *observer bias* which relates to mistakes done by the data collector and happens when the accuracy of exposure data recorded by the data collector differs systematically between cases and controls

Strengths and limitations

Strengths

- Case-control studies can be carried out rapidly and relatively cheaply compared to cohort studies.
- They are useful for studying etiology of diseases and health-related problems.
- They are useful for studying rare diseases.
- They can study multiple exposures for a single outcome.

Limitations

- They are prone to selection bias, particularly in the selection of controls.
- They are prone to information bias, because exposure status is determined after the outcome has occurred.
- They are not suitable for studying rare exposures.
- They cannot usually be used to estimate disease incidence or prevalence.