

## Epidemiological study methods

### 3.Cohort studies (Longitudinal studies)

A cohort study is an important study design especially when we search for causes and associations. The basic element of a cohort study is follow-up of a group of people without the disease for a certain period of time and watching for development of the disease (outcome) in some of them. Therefore, the starting point in a cohort study is exposure to a risk factor. We select a group of subjects who have experienced an exposure of interest, and follow them over a period of time to determine the incidence of one or more outcomes in the exposed and unexposed populations. We compare the exposed and unexposed groups in terms of the incidence of the outcome in each group.

#### **Example**

A cohort study can follow up for several years, the workers of a factory who are exposed to asbestos and determine the incidence of lung cancer among those exposed to asbestos. The study can simultaneously follow up a similar group of other workers who are not exposed to asbestos and compare the incidence of lung cancer between the two groups.

#### **What do we measure in a cohort study?**

Since we follow a population who are initially free of the outcome, the measure of occurrence which we calculate in a cohort study is incidence risk (or rate). We calculate incidence of the outcome in the exposed population, we also calculate the incidence in the unexposed population and then we compare these two by calculating a measure of effect which is called risk ratio or rate ratio.

#### **Why we conduct cohort studies?**

Cohort studies are very important in health research especially when we search for causes and associations.

1. To study etiology of diseases and health-related problems: by comparing exposed and unexposed individuals in relation to development of the disease, a cohort study can identify the risk factors which are associated with the disease and provide information about possible etiology of the disease. Analytical methods can investigate the effect of several risk factors simultaneously and in this way they can provide a better understanding for the role of each risk factor.
2. To study natural history of disease: Since a cohort study follows people without the disease into the future and watches for development of the outcome, it can provide good information about the natural history of the disease i.e. how the disease evolves over time in different individuals.

3. To study rare exposures: It is not always easy to study the effect of a rare exposure i.e. risk factors which are not common in the population. Since the starting point in a cohort study is the exposure, we can investigate the effect of rare exposures using cohort studies. For example radiation is a rare exposure not appropriate to study using a survey or case-control study because we cannot easily find enough people with the exposure in a random or a small sample. However, we can study the effect of radiation by undertaking a cohort study amongst the workers of a nuclear plant.

### **Steps in undertaking a cohort study**

#### **1. Defining the study question**

The question in cohort study is about incidence and associations i.e. the association between potential risk factors and a disease. What associations we want to study? What is the disease and what are the potential risk factors which we want to investigate? For example we may want to ask “*What is the incidence and risk factors of cardiovascular disease in the population of Sulaimani city?*”

#### **2. Defining the study population**

The study population in a cohort study is the population which we follow over time. The decision as to what population to study depends on the study question and resources available. A big and dynamic population may not be easy to follow. For example if we study risk factors of death in a small town, the study population will be all the population of the town. If we study morbidity and mortality in a cement factory, the study population will be all workers of the factory. If the exposure under study is rare such as effect of asbestos or radiation, we have to choose a suitable population in which the exposure is not as rare as the general population.

#### **3. Collecting exposure data**

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. Methods of data collection include:

- 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

#### **4. Follow up and ascertainment of the outcome**

One of the most difficult parts of a cohort study is the follow-up of the population in order to ascertain (record the occurrence of) the outcome. We have to follow up the individuals to see whether and when they develop the outcome under study. The follow up must be as complete as possible i.e. we

have to try to maximize follow-up to all individuals. This is important because if we miss a lot of people, some of whom might developed the outcome which then we would miss, this leads to bias in the results. Incomplete follow up is therefore a weakness of cohort studies.

## 5. Data analysis

The main measure of occurrence in a cohort study is incidence risk and the main measure of effect (association) is risk ratio (or rate ratio). We calculate incidence risk in the exposed group and incidence risk in the unexposed group. We compare the incidence in exposed and unexposed individual and calculate a risk ratio (or rate ratio).

*Risk of outcome in the exposed group= number of people who develop the outcome in the exposed group during the study period / total number of exposed group in study*

*Risk of outcome in the unexposed group= number of people who develop the outcome in the unexposed group during the study period/ total number of unexposed group in study*

*Risk ratio=risk of outcome in exposed/ risk of outcome in the unexposed*

Example: the following table shows data from a study of suicide by self0burning in males and females of Sulaymaniyah province in 2008. Calculate risk of suicide in each group and risk ratio comparing females to males.

Exposure group	Number of suicide	Population
Exposed (female)	133	859,963
Unexposed (male)	10	848,140

Risk of suicide in females=  $133/859,963= 15.5$  per 100,100 per year

Risk of suicide in males= $10/848,140=1.2$  per 100,100 per year

Female to male risk ratio of suicide by burns =  $\frac{\text{Incidence of suicide in females}}{\text{Incidence of suicide in females}}$

Risk ratio of suicide=  $\frac{15.5 \text{ per } 100,000 \text{ per year}}{1.2 \text{ per } 100,000 \text{ per year}}$

Risk ratio= 13

This means that the risk of suicide in females was 13 times more than the risk of suicide in males. In other words, females were 13 times more likely to commit suicide by burns than males. This is effect of female gender on self-burning.

## 6. Interpretation of results and sources of bias

Bias may cause the results of the study to be useless. Before interpretation of results we must remember that there are possible sources of error in cohort studies which are:

1. Information bias: Errors could happen in the measurement of exposures that results in inaccuracy of information collected from comparison groups. Observer bias happens when the observer (data collector) makes mistakes in obtaining/ recording data and recall bias happens when the research participant makes mistakes in remembering information of his/her exposures.
2. Selection bias: Selection bias can happen if the exposed and unexposed groups are not comparable. For example if the exposure group is workers of a factory and the unexposed group is the general population, this may cause selection bias because these two groups are not comparable.
3. Loss to follow up: Loss to follow up means we have not been able to follow some subjects and we don't know what has happened to them i.e. whether they have developed the disease or not. If a lot of research subjects are lost to follow up, this will lead to selection bias especially if the loss to follow up is not similar in the exposed and unexposed groups. Loss to follow up leads to bias because the loss does not usually occur in random.

### ***Strengths and limitations of Cohort studies***

#### ***Strengths***

- Less chance for bias: the exposure is measured at the start of the study, before the outcome occurs, and so measurement of exposure is not biased by the presence or absence of the outcome.
- Cohort studies can provide data on the time course of the development of the outcome.
- More than one outcome can be examined in one study
- Rare exposures can be investigated using appropriately selected populations.

#### ***Limitations***

- Cohort studies are usually time consuming and expensive
- Cohort studies are not good for rare diseases
- Incomplete follow up is a weakness and could bias results.