Basic Concepts in Epidemiology

The study of diseases and health-related events requires understanding and consensus on several basic concepts that underpin epidemiological studies. Here we explain some of these concepts including exposures, outcomes, cases, populations and samples.

Outcome

The outcome is the disease, the event or health-related state that we are investigating. Epidemiological studies are not restricted to study of individual diseases; any condition, state or event which might be related to health can be an outcome. For example an outcome could be lung cancer, breast-feeding, road traffic accidents, hand washing etc.

Exposure

An exposure is another name for a risk factor but it has a wider meaning. A risk factor is anything that could contribute to the causation of the disease (outcome) but an exposure, is a factor which we think might contribute to the development of the outcome. This factor may or may not be a cause of the disease but we believe that it might be related to the disease and that is why we are investigating its effect. For example if we study causes of diarrhea in babies, exposures could be breast-feeding, bottle-feeding, age, gender, mother's education, source of drinking water in the family etc. because we believe these factors could be related to diarrhea. However, the results of our study may not prove that all these factors are risk factors for diarrhea.

An exposure may even protect the person from the disease (protective exposure). For example if we study an outbreak of measles in a school, being vaccinated is a protective exposure because children who are vaccinated are less likely to get the disease. Now go back to the above example of diarrhea and think which exposures could be protective.

People who have the exposure which we study are called ***exposed*** and those who do not have the exposure are called ***unexposed***. For example in a study of the effect of cigarette smoking on lung cancer, cigarette smoking is the exposure which we want to investigate because we think it might have some relation to lung cancer. People who smoke are exposed, people who do not smoke are unexposed. There might be other factors (exposures) which could also contribute lung cancer such as age, sex, family history, occupation etc. Persons having factors that could be related to the disease are called exposed for example people with positive family history are exposed and those with no family history are called unexposed. What about sex and occupation?

Since we are not talking about diseases but health-related conditions, exposures and outcomes may interchange their roles in epidemiological research. This means that a factor could be an exposure in one study but an outcome in another. It all depends on what is the objective of the study. For example in the previous example of smoking and lung cancer, smoking is the exposure and lung cancer is the outcome. However, in a study of the effect of cigarette advertising on smoking, the exposure of interest is

advertising, and cigarette smoking is the outcome. Likewise if we do a study on risk factors of diarrhea, hand washing will be an exposure but if the objective of the study is prevalence of hand washing, hand washing will be the outcome.

The case and case definition

A case in epidemiological study is any person/subject/event which we recruit for the study. For example a case could be a child with diarrhea in a study on diarrhea; a person who smokes in a study on smoking; an incident of road traffic accident in a study of road traffic accidents; a visit to the health centre in a study on patterns of PHC attendance and so on. In order to be uniform in recruiting these cases to the study, we need a standard measure i.e. a case definition.

Case definition is a set of standard criteria used to identify cases. A case definition is essential in epidemiological research in order to identify outcomes i.e. to separate cases from non-cases. A good case definition should enable people who are involved in the study to clearly decide whether a person or event would be defined as a case or not. Imagine a study on diarrhea with no clear case definition for diarrhea. What happens when surveyors want to decide whether a child had diarrhea or not? What surveyors consider as diarrhea may not be exactly what mothers consider as diarrhea. Without a case definition, some children with diarrhea may be considered as having no diarrhea and some children with no diarrhea may be considered having diarrhea. If we define diarrhea, for example, as "having three loose stools during the past 24 hours" then chances of making mistakes while assigning diarrhea to the children will be less.

Case definition is also useful when we want to compare two studies. If the case definitions are not similar, it may not be correct to compare the findings of the two studies. Consider two studies about breast cancer; 1 out of 1000 women in population A developed breast cancer and 1.2 out of 1000 women in population B developed breast cancer in the same year. Can we compare these two populations with this information alone? We can do that only if the case definition for breast cancer is the same in the two studies. Case definitions may differ in several ways; age of women included in the study, stage of cancer and method of diagnosis.

Case definitions are developed for various situations; therefore there should be some flexibility in their development. We may have several case definitions for the same outcome depending on the circumstances. For example during an outbreak of cholera, we may define any person with watery diarrhea as a *suspected* case of cholera. We use this case definition because the situation is urgent and we need to diagnose as many cases of cholera as possible because of the danger of spread and death. Then we may have other case definitions deepening on lab examinations and other diagnostics facilities (probable *case, confirmed case*).

Broadly speaking, development of case definition depends on three factors; the purpose of the study, resources available and ethical issues.

1. The purpose of the study

If the purpose of the study is to compare the results of our study to previous results of other studies, we have to use the same case definitions. There are standard case definitions for any disease. For example WHO's International Classification of Diseases 10^{th} revision (ICD-10) provides standard labeling for most diseases. If the purpose of the study is to investigate the risk factors of a disease, we should use precise case definitions. For example if we want to investigate the effectiveness of measles vaccine, we have to use a precise case definition for measles based on serology as clinical diagnosis may not be very accurate. On the other hand, if we want to quickly investigate an outbreak of cholera, where there is urgency to control the epidemic, we may use loose case definitions based on most common clinical features.

2. The resources available

If we have only limited resources available (such as financial, human etc.) we may not be able to use expensive instruments or diagnostic tests to identify cases. In such a situation we may prefer to use a case definition based on clinical diagnosis needing less tests ad resources. However, we must still use clearly defined criteria and use them for diagnosis of all cases. The WHO has developed a set of simple clinical criteria for diagnosing certain disease such as TB, lower respiratory tract infection, HIV/AIDS and others which are often used for epidemiological surveillance in countries with limited resources.

3. Ethical Issues

We have always to respect rights of the study subjects (people we study) and ensure their safety. Criteria used in a case definition must not be harmful to the perosn. For example invasive techniques are not feasible for case definitions of population surveys because they may be harmful to people in the study. Where there are chances of harm to people, use of these invasive techniques are considered non-ethical or ethically not acceptable. Invasive techniques may be feasible for case definition of hospitalized patients but in population surveys and other studies it will be more feasible to use case definitions based on non-invasive criteria.

The population

In epidemiology population is not restricted to people living in a geographical area. Population could be any group of research subjects whom we want to study. For example the population could be all population of Sulaimaniyah in one study; it could be only children under 5 years of age in another study; or it could be pregnant women of local area.

Before doing any study we have to define the target population and the study population clearly. The ***target population*** is the larger population on which you would like to generalize the results of your study and therefore which includes the study population.

The ***study population*** is part of the target population from which we select people who participate in the study. The study population is more limited in number and more accessible for the researcher than the target population. For example in a study on anemia during pregnancy at a Primary Health Centre, the target population is all pregnant women at the catchment area of the PHC and the study population could be all pregnant women who are registered at the PHC.

The sample

In epidemiological studies we usually do not have the capacity and resources to study the entire population and collect information about all individuals in the population. Instead we collect data from a much smaller group of the population which is called a ***sample***. A sample therefore is part of the target population or study population. In order to be a good example of the bigger population, a sample should be representative to the population from which it is drawn i.e. the characteristics of the subjects in the sample should reflect characteristics of the target population. If a sample is not representative to the target population, we cannot generalize findings of our study to the target population. There are many ways to select a representative sample, but the basic principle is chance. Chance should decide who will be included in the sample. If the investigator rather than chance decides who is included in the study, the results will not be correct.