

Epidemiological study methods

1) Cross-sectional studies (Prevalence studies/ Surveys)

Cross-sectional surveys are study methodologies that investigate the population at a point in time by collecting information on exposures and outcome at the same time. In a cross-sectional study, we measure the frequency of a particular *exposure(s)* and / or *outcome(s)* in a defined population *at a particular point in time*. As cross-sectional studies collect data on existing (prevalent) cases, they are also called prevalence studies.

For example, in an survey on smoking, we collect information on the smoking habits of each individual (outcome) in a population. We can also collect information on some potential risk factors, such as age, sex, occupation, education of the individual (exposures). Our primary aim in this study would be calculation of the prevalence of smoking.

Why we conduct surveys?

Surveys are one of the most common methods of health research. They are relatively quick and easy to undertake and provide important information about the health situation which is essential for planning, providing and improving health services.

1) Health service planning: one of the most important uses of surveys is to collect information on the various diseases, states and health-related problems in the society. Such information is essential for health authorities in order to plan health services and deliver them appropriately and as needed. Health policy must be based on evidence and evidence could be collected from surveys and other research methods.

Examples: A survey can collect data on the utilization of services and this information can be used for improving service delivery. Some countries undertake large scale health and demographic surveys (e.g. MICS in Iraq) from time to time to obtain health-related data from a random sample of the population in order to understand the health status needs of the population. We can also use surveys to collect information of health related attitudes and behaviors such as smoking, exercise and diet which are essential to plan interventions to tackle these issues.

2) Evaluation of health programs: Surveys can provide baseline information about a health problem before an intervention programme is introduced. When the programme is implemented to address the problem, another survey can collect information and compare the results with the previous one, and in this way the effectiveness of the programme can be evaluated. For example, if health authorities want to know whether the EPI vaccination programme has been effective in reaching a good proportion of children, they undertake vaccination coverage surveys to evaluate their programme.

3) Generate hypothesis about etiology: by calculating prevalence and prevalence ratios in different groups of population, analysis of cross-sectional studies can help generate

hypothesis about possible causes of diseases and provide quick initial information on etiology. For example the observation that coronary heart disease is more prevalent in men than in women has led to hypotheses about the possible protective effects of estrogens.

Steps in undertaking a survey

1. *Defining the study question*

What kind of information we want to collect? What is the disease or health-related problem which we want to investigate? The question in surveys is about prevalence i.e. the frequency of occurrence of the exposures and outcomes in the population during the time of the survey. For example we may want to ask “*What is the prevalence of diabetes in Sulaimani population?*”; “*How are the blood groups distributed in the Kurdish population?*”

2. *Defining the target and study populations*

It is very important to define the target population and the study population. The target population is the larger population on which you would like to generalize the results of your study and therefore which includes your study population. The study population is part of the target population from which we draw the sample. The study population is more limited in number and more accessible for the research than the target population. In a study of childhood anemia in Sulaymani city, the target population will be all children of Sulaymani city and the study population could be children who visit the health centres for any reason.

3. *Sample selection*

When we have defined the study population we usually need to take a sample of this population. This is because usually the study population is large and we will not have resources and time to collect information from every individual. Therefore we only collect information on a sample of this population. Which people we collect data from? This sample should be representative to the population and one way to ensure that, is to take a random sample of the population. There are established methods to calculate the number of individuals we have to include in the study (sample size) and there are also specific methods for selecting individuals to be included in the sample (sample selection). Sample size calculation and sample selection methods are essential parts of any health research project.

4. *Data collection*

Before starting the study, we have to define what data collection tools we use. We may use standard tools which have been present before and used by other researchers. This will be an advantage if it applies to our study. Frequently we have to develop our own data collection tools such as questionnaires. We must use standard procedures for data collection to make sure that the same information is collected from all individuals in the same way. Main data collection tools in health research include the following:

- Personal interviews: face-to-face interview with individuals using a questionnaire or Telephone interviews using a questionnaire

- Self administered questionnaire where the questionnaire is delivered by post or personally to the person.
- Medical records: previous medical records of the individuals
- Physical examinations
- Diagnostic tests

5. *Data analysis*

When all data have been collected, they are usually entered into a computer programme, cleaned and analyzed. Data analysis depends on the type of data collected and objectives of the study. In surveys, we calculate distribution of the exposures and outcomes in the population and in various sub-groups of the population. The following data is from a survey of anemia in women. The researcher wants to know whether literacy is related to anemia. We first calculate overall prevalence of anemia in the sample, then prevalence of anemia in literate and illiterate women separately and finally we calculate prevalence ratio comparing illiterate to literate women. .

	All women	Women with anaemia
Literate	250	50
Illiterate	100	50
All	350	100

Overall prevalence of anemia= number of women who are anemic/ total number of women in the study= $100/350=0.29=29\%$

Prevalence of anemia in the exposed (illiterate women) = number of illiterate women who are anemic/ total number of illiterate women in the study= $50/100=0.50=50\%$

Prevalence in the unexposed (literate women)= number of literate women who are anemic/ total number of literate women in the study= $50/250=0.20=20\%$

Prevalence ratio of anemia in illiterate compared to literate women= prevalence of anemia among the illiterate women/ Prevalence of anemia among the literate women= $0.50/0.2=2.5$

What does this mean?

6. *Interpretation of results*

Before interpretation of results we must remember that there are possible sources of error which we should be aware of. We should also be aware of strengths and limitations of surveys while interpreting the results. If the sample selection is not random the results will probably be biased i.e. different from the truth. Bias is any error in the design and conduct of a study that makes the results different from the truth.

- Selection bias happens when the sample is not representative to the population and therefore the groups which we want to compare are not well-represented in the sample.

- Reporting bias is a form of information bias and happens when the participants do not provide the true information for any reason. Sometimes they forget the true information “recall bias” for example in a vaccination survey a mother may say that her child is vaccinated while he is not. If inaccurate information was provided then our results will be biased.
- Observer bias is another form of information bias where people who collect the information (data collectors) make mistakes in obtaining/ recording the data. If these people make plenty of mistakes then our results will be biased.

Apart from sources of bias mentioned above, we have to remember that in a cross-sectional survey, we collect data on exposure and outcome at the same point in time. Because of this, it will be difficult to tell whether the exposure occurred before or after the outcome and therefore decision about etiology is not possible.

Strengths and Weaknesses of Cross-sectional Studies

Strengths

- Surveys are relatively quick, easy and economical to conduct
- They provide important information on the distribution and burden of disease in populations
- They can provide initial information on exposure-outcome relationship and hence can provide initial hypothesis about etiology

Limitations

- Surveys measure prevalent cases rather than incident cases. Because of this, they are of limited value for investigating etiological relationships between risk factors and outcomes.
- Since information is collected on the exposures and outcomes at the same time, it is difficult to establish the time-sequence of events