UNIT 5 ISSUES IN THE DESIGN AND CONDUCT OF SELECTED EPIDEMIOLOGICAL RESEARCH DESIGNS

Structure

5.1 INTRODUCTION

Unit 3 and 4 presented an orientation to the descriptive and analytic research designs used in epidemiology. Designing and conducting these research design requires basic understanding of the characteristics of these designs. In this unit we will briefly review these characteristics and deal mainly with the steps involved in the conduct of these research designs.

Objectives

After studying this unit, you will be able to:

- explain the characteristics and features of types of descriptive research such as correlational studies and case-control studies,
- describe the steps involved in descriptive research,
- elaborate on the issues involved in the design and conduct observational studies,
- enumerate the steps involved in experimental research, and
- identify and explain a few designs for experimental studies.

5.2 DESCRIPTIVE RESEARCH

Descriptive research studies, as you may recall studying earlier in Unit 3, are designed to obtain information concerning the current status of a given phenomenon. They are concerned with the existing conditions or relationships, prevailing practices, current beliefs, points of view or attitudes, processes that are going on and their effects and the developing trends. In short, it determines the nature of a situation as it exists at the time of study. The aim of descriptive research is to describe "what exists" with respect to variables or conditions in a situation.

The descriptive research method is appropriate in behavioural sciences. Many types of behaviour that interest the researcher cannot be arranged in a realistic setting. For example, it would be unthinkable to prescribe cigarette smoking for the purpose of studying its possible relationship to throat or lung cancer, or deliberately arrange accidents, in order to evaluate the effectiveness of seat-belts or helmets in preventing serious injuries.

Although some experimental studies of human behaviour can be appropriately carried out both; in the laboratory and in the field, the prevailing method used in social sciences is descriptive. Under the conditions that naturally occur at home, human behaviour can be systematically examined and analyzed. This analysis may lead to the modification of factors or influences that determine the nature of human interaction. It is through this modification of factors that social institutions may become more effective influences in promoting human welfare. In this context, here in this section we shall look at issues to be considered in the design and conduct of descriptive studies.

5.2.1 Descriptive Research: Main Steps

In descriptive studies, we do not present private convictions and data based on casual or cursory observations. In a descriptive study, we

i) examine the problematic situation,

ii) define our problem and state our hypothesis,

iii) list the assumptions upon which our hypothesis and procedures are based,

iv) select appropriate subjects and source materials,

v) select or construct techniques for collecting data,

vi) validate the data gathering techniques,

vii) make objective and discriminating observations,

viii) describe, analyze and interpret our data in clear, precise terms.

We, as researchers, collect evidence on the basis of some hypothesis, tabulate and summarize the data carefully, and then analyze the results thoroughly in an endeavour to draw meaningful generalizations that will advance knowledge. Let us understand this process.

Collection of data

When presenting a descriptive research report, one must identify not only the kind of data obtained but also the exact nature of its population. The units that constitute a population may be people, items, events or objects. After identifying the population, one must decide whether to collect data from (a) the total population or (b) a representative sample of the population. Let us consider.

a) *Total Population*: Obtaining information from every unit of a small population is not difficult in most instances, but the findings are not applicable to any population other than the group studied. After collecting information from every student in one particular study center, for example, you may draw generalizations about the average age or the kind of profession of students in the center, but, you cannot claim that these generalizations will hold true for students in any other study center. Similarly, after studying the attitude of mothers towards breast feeding in an urban setting, a researcher cannot claim that the findings will hold true for mothers in other settings.

b) *Sample Population*: Obtaining information from a large population, such as all mothers in the state, is often impractical, impossible or exorbitantly costly. Contacting, observing, measuring or interviewing every unit in the group may

absorb so much time that the data becomes obsolete before the study is complete. To overcome these difficulties, investigators often collect information from a few carefully selected units drawn from a population. While carrying out the research on eating habits of adoloscents of Delhi, the researcher cannot collect data from all the adoloscent living in Delhi. He/she will have to select a representative sample of the population. However, the sample will have to be selected from different categories of the population, viz., male/female, rural/urban/slum, employed/ unemployed etc. If these sample units represent accurately the characteristics of the population, generalizations based on the data obtained from them may be applied to the entire group. But selecting a representative sample has already been discussed earlier in Unit 2.

Descriptive data may be expressed qualitatively (in verbal symbols), as well as, quantitatively in mathematical symbols. A study may consist almost exclusively of one form or may contain both forms. If an investigator is comparing the nutrition education programmes of selected anganwadis, he/she may present qualitative data, i.e., word descriptions concerning the organization of the programmes, the duties and competencies of the anganwadi workers, supervisors. But he/she may also include considerable quantitative data, such as the amount of time spent by workers, if any on development of the printed material, the number of beneficieries who attended the meeting and the amount of specialized training which the workers have received.

Qualitative data – word description – may predominate in studies that examine the general nature of a phenomena. Qualitative studies give social scientists useful information, but verbal symbols lack precision as words do not hold the same meaning for all people, at all times, and in all contexts. However, qualitative studies need not be looked down upon; for, they help workers identify the significant factors to measure. Until these general explorations are made, measurement cannot be utilized fruitfully.

With this basic discussion regarding the general issues basic to descriptive studies, let us move on to the specific issues linked with the design and conduct of different types of descriptive studies. We began with correlation studies.

5.2.2 Correlation Studies: Basic Issues

Human behaviour at both, the individual and the social levels is characterized by great complexity. However, given the present state of social research, we understand too little of this complexity. One approach to a fuller understanding of human behaviour/ practice is to begin by testing out simple relationships between those factors and elements which are supposed to have some bearing on the phenomenon in question. The value of correlational research is that it is able to achieve this end. We know that one of the primary purposes of science, as conceived traditionally, is to discover relationship among phenomena with a view ultimately to predicting and, in some situations, controlling their occurrence.

Much of social sciences research in general, and nutrition/health research in particular, is concerned at our present stage of development with the first step in this sequence, i.e., establishing interrelationship among variables. Correlational studies are concerned with determining the extent of relationship existing between variables. They enable us to measure the extent to which variations in one variable are associated with variations in another. We may wish to know, for example, how delinquency is related to social and class background, or whether a relationship exists between the dietary components and coronory heart disease or there is a link between maternal anaemia and birth weight.

Correlational studies are generally intended to answer three questions. They are:

a) Is there a relationship between two variables (or two sets of data)? If the answer to this question is 'yes', then two other questions follow

b) What is the direction of the relationship? and

c) What is the magnitude of the relationship? The magnitude of the relationship is determined by the coefficient of correlation.

For instance, on the basis of his/her experience, a researcher may hypothesize that there is a relationship between performance in an intelligence test and a test of achievement in arithmetic. The correlational technique will help him test his/her hypothesis about the relationship. Pearson's product moment, one of the best known measures of association, is a statistical value of the coefficient of correlation ranging from −1.0 to +1.0, through zero and expresses a relationship in quantitative form. Where the two variables fluctuate in the same direction, i.e., as one increases so does the other, a positive relationship is said to exist. A negative correlation or relationship, on the other hand, is to be found when an increase in one variable is accompanied by a decrease in the other variable. The values near zero indicate a weak relationship between the variables, whereas values closer to either +1.0 or −1.0 indicate a stronger relationship in either of the directions. Thus, the coefficient of correlation, tells us something about the relationship between *two variables*. However, other measures exist which allow us to specify relationships when more than two variables are involved. These are known as measures of *multiple correlation* and *partial correlation*. (We will not go into details about these measures over here).

One danger in interpreting correlations is to assume that because two variables are related in a predictable fashion to one another with a high degree of probability, they are also in a causal relationship. This is not necessarily the case. For one thing, there is never more than a probable relationship between variables in any case. For another, it is quite possible for two variables to be related to one another with a high degree of probability but with a third variable accounting for the nature of relationship. Correlation must not be interpreted to mean that one variable is causing the scores in the other variable to be what they are. For example, it may be found that there is a negative correlation between measures of anxiety and measures of intelligence. It should not be interpreted that there is a causative relationship between anxiety and intelligence, that is, that pupils are anxious because they are unintelligent. It might be that there are other underlying characteristics of individuals that tend to make some appear unintelligent and anxious, and others, intelligent and not anxious. Interpretation of such a correlation is difficult without experimental confirmation. For example, the relationship between anxiety measures and intelligence measures could be investigated experimentally by deliberately inducting anxiety in a testing situation and determining the effect on intelligence test scores.

Next, let us consider the issues in the design of case study method.

5.2.3 Case Study Method

Case studies may include the study of different individual units like the family, an organization/institution, a disease condition or an nutritional/health social welfare programme. (Please note the meaning of the word 'unit' in this context).

In the case of studies on communities, a village, a tribe, a slum area or a particular group of people, each can be considered a unit of investigation.

Whatever the unit of a case study, it is treated as a whole in the context of specific situations. The wholeness is determined through an abstraction of ideas. In one case, an individual's specific behaviour may be perceived as a totality; in another case, a situation consisting of group activities may be treated as a whole. Especially, in nutrition educational situations, the units under investigation could be a whole instructional programme, a micro-instructional system, instructional development in a 'group setting' or in an 'individual setting', allowing the possibility of using a single method or integration of a number of methods.

Before we move on to the steps involved in the design of case-studies, let us recapitulate the purpose and basic characteristics of case-studies.

Purposes of Case Studies

Usually, case studies are conducted for developing a deeper understanding about intricate relationship existing in the process-aspects of a specific unit/units through qualitative investigations as discussed earlier in Unit 3. In this context, the case study method is not very different from the approaches of naturalists. So, many a time, the case study method is treated as a kind of naturalistic inquiry.

Case studies are conducted with a clinical purpose. They are treated as diagnostic and prognostic measures for clients' treatment. This approach has a psycho-therapeutic background.

Characteristics of Case Study Method

The procedural aspects of a full-fledged case study display certain specific characteristics, viz., continuity in investigation, completeness, authenticity of data, confidential recording, and intellectual syntesis. We shall explain each one of them briefly as follows:

i) *Continuity in Investigation*: Continuous and prolonged enquiry about the situations is necessary till the underlying factors are explored and plausible patterns of their interaction relationship identified. For example, the problems underlying the failure of a nutrition education programme cannot be explored in one go. A researcher may have to undertake prolonged inquiries.

ii) *Completeness*: A sound case study involves extensive collection of data concerning internal, as well as, external environment of the unit under study. Data collection continues till the completeness of data is ensured and a complete picture of the unit emerges.

iii) *Authenticity of Data*: A report of the case study must be based on meaningful, reliable and valid information regarding the case. Several qualitative and quantitative techniques such as interviews, observations, record surveys and administration of test questionnaires find their appropriate application in case studies. Use of multi-techniques approach to data collection and cross-examination of data through different techniques can take care of the authenticity of data. In a case study on Integrated Child Development Services (ICDS) for example, the problems faced by beneficiaries will have to be cross-checked with the functionaries in the system, as well as, with the existing records. Moreover, since the researcher interacts with the typical situations personally, most of the ethical issues regarding the nature of data, the sample situations, or sample respondents, the nature of interactions etc., emerge during the investigation, these issues need to be dealt with care to make the case study ethically meaningful.

iv) *Confidential Recording*: The necessary data, involving personal and ethical issues like relationships of anganwadi workers with the benficiaries, confidential records, documents about the institution etc., must be handled tactfully and every care must be taken to maintain their secrecy.

v) *Intellectual Synthesis*: Since a case study involves multimethod inquiry and deals with all significant situations concerning the unit, appropriate synthesis of the data is necessary to depict the uniqueness of the unit and to explore significant relationships. A skilled investigator with theoretical sophistication, insightfulness and writing skills can do justice and prepare a sound case study.

Now with the basic understanding about the characteristics of case study, let us study the steps involved in the conduct of case studies.

Case Study: Main Steps

Most researchers treat case study method as one of the forms of naturalistic inquiry. Therefore, the case study method follows the same steps as are followed in the case of naturalistic inquiry. However, the following six steps are considered to be very significant.

i) *Selection of a Case for Investigation*: The first step in any case study is the identification and selection of a case for investigation. It mostly depends on the basic questions of the researcher, such as: 'am I interested in the phenomenon? Or am I determined to solve the problems of a typical institution? Or am I interested in identifying the underlying factors contributing to the excellent performance of an institution? Once the case is identified, then one needs to determine the status of the case. For this, several pieces of preliminary information are collected about the background of the case through the already available sources. At this stage, the initial exercise in setting the course of research is done; it comprises the following:

- Demarcation of the relevant aspects of the case to be investigated,
- Preparation of a broad outline of the study of sample situations, and
- Preparation of the appropriate tools for collecting the 'benchmark' data about all the pertinent aspects of the case under study.

Answers to these questions would tell us whether we have identified the 'case' or not.

ii) *Data Collection*: Now we move on to the stage of data collection. In the process of collecting benchmark data about the case, we may make use of both qualitative, as well as, quantitiative techniques like observations, interviews, check lists, proforma, open-ended questionnaires, record surveys, clinical tests, etc. Every care must be taken to use the tools specifically relevant for the case. In most cases, the first round exploratory work is done through personal interaction with the situation under study.

iii) *Analysis of First Round Data*: Through systematic analysis of the first round data, we can identify the more complicated situations or problems, and raise pertinent questions about the influential factors. In the case of clinical investigations, we can state various hypotheses about the solutions to the problems.

iv) *Second Round Investigation*: The second round investigation is conducted for only those specific questions or factors which are identified through the analysis of the first round data. Intensive investigations about these specific issues/problems are conducted through prolonged observations, informal and formal interviews, questionnaires, cross-examination of different documents and records, administration of specific tests etc. At the end of this second round of data collection, analysis and interpretation of data begin. However, during the interpretation of the data, if some more evidence is needed we may go for another round of data collection. Actually, in a case study, the process of data collection, its analysis and interpretation go on in a cyclical order till satisfactory answers to the questions arising in the course of investigation are found and a clear cut picture of the case emerges through investigation.

v) *Introduction of Alternative Measures*: In the case of clinical studies, the most suitable alternatives as hypothesized through investigations are introduced at the fifth stage. For example, after a case study on a disease condition, problems may be identified and suggestions for improvement made. These remedial measures may be tried out and their effect observed with respect to improvement in the disease condition.

vi) *Follow-up Activities*: Investigations should be made regarding the effectiveness of the alternative measures introduced. Such investigations give us feedback on the strengths and weaknesses of the corrective measures. If we find them to be less effective, we should conduct further studies to arrive at some 'newer' remedial measures and apply them to the case.

With this basic review of the steps, we hope you would be able to design and conduct case study as part of your research.

Given next is a questions to help you recapitulate what you have learnt so far. Answer this questions and then move on to the study of issues involved in the design and conduct of observational studies.

Check Your Progress Exercise 1

1) Explain briefly the purpose of correlational studies.

.....

.....

.....

.....

.....

2) What steps would you follow in conducting a case study.

.....

.....

.....

.....

.....

5.3 OBSERVATIONAL STUDIES

Observational studies as already mentioned earlier in Unit 4 are epidemiological studies in which observations are made but investigators do not control the exposure or intervention and other factors. Cohort, case-control and cross-sectional studies are the main observational studies. Here in this section we shall briefly review the issues involved with the design and conduct of cohort and case-control studies.

5.3.1 Issues in the Design of Case-Control Studies

Case control studies as already described earlier in Unit 4 are observational analytic investigations which provide for evaluation of the association between an exposure and a disease. For a case-control study to be able to provide sound evidence of a valid statistical association, comparability of cases and control is essential. In this context, therefore, major issues to be considered in the design and conduct of case-control studies focus on a) definition of disease or outcome of interest, b) selection of cases, and c) source of information and ascertainment of disease and exposure status. Let us review these issues in greater details.

A. *Defining the Disease or Outcome of Interest*: The first issue to be considered in the design of case-control study is to define the disease or outcome of interest. It is important that we define the disease/outcome as far as possible as a homogeneous entity since very often, you will find that similar manifestational entities of disease have very different etiologies. Therefore to be able to define the disease/outcome clearly it is imperative that we establish strict diagnostic criteria for the disease. For example, case-control studies on anaemia often use diagnostic criteria formulated by the World Health Organization (WHO).

B. *Selection of Cases*: Once the disease/outcome has been defined based on a standard diagnostic criterion, it is easy to identify cases with the disease under consideration. Cases can be selected from a number of sources. These may include medical records from hospital/clinics during a specified period of time or

selecting subjects with the disease in a defined general population or random sample from a defined population at a single point or during a given period of time. The first approach is referred to as a *hospital-based case-control study* and the second, the *population-based case-control study*. Regardless of the source, the cases selected for the study can represent either incident (newly diagnosed) or prevalent (existing at a point in time) cases of the disease. Once the cases have been identified, next issue deals with comparable controls for the study.

C. *Selection of Controls*: This step you would realize is the most difficult and perhaps critical issue in the design of the case-control study. There is a need to obtain comparable information from cases and controls. Controls are necessary to allow the evaluation of whether the frequency of an exposure or a specified characteristic observed in the case group is different from that which would have been expected based on the study of a series of comparable individuals who do not have the disease. As in the cases, several possible sources of selecting controls include hospital controls (consist of patients at the same hospital, who have been admitted for conditions other than the disease being studied), general population controls (selected through canvassing households in targeted neighbourhood, random-digit telephone dialing or from voting list, population register etc.) and special control series such as friends, neighbours or relative of cases. Each offers particular advantages and disadvantages that must be considered for any particular study. For example, the important practical and important advantage of using hospital controls is that they are easily identified and readily available in sufficient numbers thus minimizing the cost and effort involved in their assembly. Further, they share the same selective process by which the cases were identified. The chief disadvantage of hospital controls, however, is that they are by definition ill (since they may be suffering from another illness) and therefore differ from healthy individuals in a number of ways that may be associated with illness or hospitalization in general. Similarly the use of general population controls assures the greatest comparability, since they come from the same source population that give rise to the cases. But identifying and interviewing population controls is more costly and time consuming and also the quality of the information may differ between the cases and control because those in the general population may not recall exposures with the same level of accuracy as those who have developed the disease. Once the source of the control series has been determined, we may need to consider another issue that is to include one, two or more control groups. Ideally a single control group is most suited for the case-control study.

D. *Deciding the Number and Manner of Selection of Cases and Control*: Once the controls have been determined, it is important to decide how many and the manner in which individual subjects will be selected. If the available cases and controls is large and the cost of obtaining information from both groups is comparable, then the optimum case-to-control ratio is 1:1. When the sample size of cases is limited and also the cost of obtaining the information from the cases is greater, the control-to-case ratio can be altered to achieve the desired sample size. Note, as the number of controls per case increases the power of the study also increases. It is not generally recommended that that this ratio increase beyond 4:1.

E. *Ascertainment of Disease and Exposure Status*: Once we have defined the case and the control series, the next issued deals with obtaining the information on the disease and exposure. Information about the disease status can be obtained from many sources including review of death certificate, case registers that maintain on-going surveillance, medical record of physicians/doctors/hospitals etc., hospital admission or discharge records etc. Information about exposure can be obtained from the subject themselves, by either interview or mail questionnaire, from a surrogate, such as spouse of the participants or caretakers of children or from information records in medical records.

One of the central issues in the ascertainment of the exposure in a case-control study is the basis on which a given individual should be considered exposed. This decision will involve defining the part of the person's exposure history that is considered relevant to the etiology of the disease under study and will depend on the understanding of the mechanism of the disease process.

We hope having gone through the issues highlighted above you may have got a good insight into how to design and conduct case-control studies. Next, we shall focus on cohort studies.

5.3.2 Issues in the Design of Cohort Studies

Cohort studies we know are those observational analytic studies which observe a large group of people over a period of time. Refer to Figure 4.3 in Unit 4 which illustrates the cohort design. It must be evident that in such a design the groups of person (cohort) to be studied are defined in terms of characteristics manifest prior to appearance of disease under study, and these defined study groups are observed over a period of time to determine and compare the frequency of study disease among them. Major issues to be considered in the design and conduct of cohort studies focus on selection of exposed population, selection of comparison group, source of data and ascertainment of outcome data through follow-up. Let us consider these issues.

A. *Selection of Exposed Population*: Which particular group to be selected to comprise the exposed population in a cohort study is the first issue? Your choice of a particular group will depend on frequency of the exposure under consideration, the need to obtain the complete and accurate exposure and follow-up information on all study subjects, the nature of problem/exposure or research question being evaluated, hypothesis under investigation and specific features of the design. For example, for relatively common exposures, such as cigarette smoking or coffee drinking, a sufficient large number of exposed individuals could possibly be identified from a number of possible populations. For rare exposures, it would be more efficient to choose a group especially because they have undergone some unusual exposure or experience, the effects of which are yet to be completely evaluated. To illustrate selecting individuals living near a suspected environmental hazard, such as nuclear testing ground or toxic waste dump site or industrial unit etc. To allow for the investigation of a number of common risk factors for relatively common chronic diseases, the best choice of study population might be a general cohort, drawn from a geographically and demographically well-defined area, which can be surveyed to establish base-line exposure status with respect to a number of factors and then examined periodically to ascertain future outcomes as in the case of the Framingham Heart Study about which you may recall studying in Unit 4. The use of these cohorts can lead to identification of etiological agents in the circumstances being investigated. The advantage of selecting a specific exposure population is that it allows the accrual of sufficient exposed individuals in a reasonable period of time.

B. *Selection of Comparison Group*: Once we have identified the source of the exposed group, the next consideration is the selection of an appropriate comparison group or non-exposed individuals. The major principle underlying this decision should be that the groups being compared should be as similar as possible with respect to all other factors that may be related to the disease except the factor/ variable under consideration, so that if there is really no association between the exposure and the disease, the disease rates in the populations being compared will be essentially the same.

C. *Sources of Data*: While designing the cohort study we would also need to consider the availability of accurate and complete information regarding the subjects that will allow the classification of all cohort members according to whether they

have been exposed to the factor(s)/variable(s) under investigation or developed any of the outcomes of interest. Information related to exposure can be obtained from records collected independently for the study or from medical and employment records, information supplied by the study subject themselves, data obtained by medical examination or other testing of the participants or direct measurement of the environment in which the cohort members have lived or worked. Outcome information may also be ascertained from existing records, medical records, medical/physical examination, from questionnaires or from death certificates.

D. *Follow-up of Study Subjects*: Finally, the most crucial consideration in the cohort research design is the ascertainment of the outcome data which would involve tracing or following all study participants from the point of exposure into the future, to determine whether they develop the disease of interest. Outcomes you would realize can have a latency period of days to weeks (such as acute illnesses) to months (such as congenital malformations) or many years decades (such as degenerative diseases, cancer etc.). The longer the observation period required the more difficult it would be to achieve complete follow-up, because people are more likely to move, change jobs, loose touch with the study group etc. Thus you would realize that collecting follow-up data on every person enrolled represent a major challenge in the design of a cohort study.

We hope the issues highlighted above would guide you adequately in designing and planning a cohort study as part of your research. We shall end our study of the issues involved with the design and conduct of observational studies here.

Next, we shall focus on experimental research and experimental study designs.

5.4 EXPERIMENTAL RESEARCH

Experimental research studies, as you may already be aware, are designed for establishing causal relationships. This method begins with a question concerning the relationship between two or more variables. At the same time, the researcher advances one or more hypotheses stating the nature of the expected relationship. The experiment is the event planned and carried out by the researcher to gather evidence relevant to the hypotheses.

The application of experimental method yielded better results in physical sciences; therefore, this method was soon applied to other sciences like biological sciences and medicine.

In its simplest form an experimental has three characteristics:

i) an independent variable is manipulated,

ii) all other variables except the independent variable are held constant, and

iii) the effect of the manipulation of the independent variable on the dependent variable is observed.

The independent variable and the dependent variable(s) are important in an experiment. The independent variable is manipulated or changed by the experimenter. The variable upon which the effects of changes are observed is called the dependent variable, which is observed but not manipulated by the experimenter. The dependent variable is so named because its value is hypothesized to depend upon, and vary with, the value of the independent variable. For example, to examine the effect of different strategies to prevent anaemia, an investigator would manipulate method, the independent variable, by using different strategies (namely, supplementation nutrition education, fortification etc.) in order to ascertain their effect upon anaemia, the dependent variable.

The basic characteristics of experimental research are described next.

5.4.1 Three Characteristics of Experimental Research

There are three essential ingredients in the conduct of an experiment: control, manipulation and observation. We shall discuss each of them as follows:

i) *Control*: Control is the first essential ingredient of experimental method. Without control, it is impossible to evaluate unambiguously the effect of an independent variable. Basically, the experimental method rests upon two assumptions regarding variables. These are:

   a) If two situations are equal in every respect except for a variable that is added to or deleted from one of the situations, any difference appearing between the two situations can be attributed to that variable. The statement is called the *law of the single variable.*

   b) If two situations are not equal, and it can be demonstrated that none of the variables is significant in producing the phenomenon under investigation, or if significant variables are made equal, then any difference occurring between the two situations after the introduction of a new variable to one of them can be attributed to the new variable. This statement is called *the law of the only significant variable.*

   The main purpose of 'control' in an experiment is to arrange a situation in which the effect of variables can be measured. The conditions to be fulfilled under the first law can be obtained more easily in physical sciences. A high degree of control is much easier to achieve in a laboratory setting than in situations outside the laboratory. In the laboratory, there is only a limited number of variables which can be manipulated easily. However, as epidemiological research is concerned with distribution and determinants of disease, there are always many variables present in a situation. 'Control' is the term used to indicate in experimenter's 'procedures' for eliminating the differential effects of all variables extraneous to the purpose of the study. (An extraneous variable is a variable that is not related to the purpose of the study but may affect the dependent variable). The experimenter exercises controls, for instance, when the groups are made comparable on extraneous variables that are related to the dependent variable. If a variable is known to be unrelated to the dependent variable, it cannot influence the dependent variable and we do not need to control it for its effects.

ii) *Manipulation*: Manipulation of a variable is another distinguishing characteristic of experimental research. It refers to a deliberate operation performed by the researcher. In contrast to the descriptive research, in which the researcher simply observes conditions as they occur naturally, the researcher in the experimental research actually sets the stage for the occurrence of the factors whose performance is to be studied under conditions where all other factors are controlled or eliminated. In epidemiological research and other behavioural sciences, the manipulation of a variable takes a characteristic form in which the experimenter imposes a predetermined set of varied conditions on the subjects. This set of varied conditions is referred to as the *independent variable*, the *experimental variable*, or the *treatment variable*. Then, different conditions are designed to represent two or more values of the independent variable; these may be differences in degree or differences in kind. That is, the independent variable may assure two or more values and the difference in the values may be of quantitative or qualitative nature. Methods of treatments, attitudes, socio-economic status, personality characteristics, etc. are some common examples of the independent variable in epidemiological research. For example, if the researcher compares two methods of treatment, then method of treatment is the independent variable and can be manipulated by the teacher. We may manipulate a single variable or a number of variables simultaneously.

iii) *Observation*: In experimentation, we are interested in the effect of the manipulation of the independent variable on a dependent variable. Observations are made with

respect to some characteristics of the behaviour of the subjects employed in the research. These observations, which are quantitative in nature, may constitute the dependent variable.

With this basic review of the characteristics let us next get to know about the steps involved in experimental research.

5.4.2 Steps Involved in Experimental Research

A number of steps are involved in experimental research. Here, we shall talk about four steps to reach the stage of the 'actual experiment'. Brief explanations are needed for steps 3 and 4 only. The steps are

i) Surveying the literature related to the problems,

ii) Identifying and defining the problem,

iii) Formulating a problem hypothesis, and defining basic terms and variables. Stating of hypotheses is an important step in experimental research. They suggest that an antecedent condition or phenomenon (independent variable) is related to the occurrence of another condition, phenomenon, event, or effect (dependent variable). To test a hypothesis, the experimenter attempts to control all the conditions except the independent variable which he/she manipulates. Then he/she observes the effect on the dependent variable presumably because of the exposure to the independent variable.

iv) Constructing an experimental plan is the next step in experimental research. This refers to the conceptual framework within which the experiment is conducted. This would involve:

- Identifying all the non-experimental variables that might contaminate the experiment and determine how to control them. If the researcher is interested in finding out the effect of two particular teaching methodologies on achievement in arithmetic, the pure effect of teaching methodology may get contaminated if the student gets extra coaching in arithmetic at home or some competition is held in the school,
- Select a research design,
- Select a sample of subjects to represent a given population, assign subjects to groups, and assign experimental treatments to the groups. This aspect has already been discussed earlier too in Unit 4 under the sub-section 4.4.1.
- Select or construct and validate instruments to measure the outcomes of the experiment,
- Outline procedures for collecting the data and possibly conduct a pilot or "trail run" test to perfect the instruments or design, and
- State the statistical or null hypothesis.

The above steps bring the researcher to the stage when he/she actually conducts the experiment, applies statistical measures to the data obtained, and then tests the significance of the results.

In the next sub-section, we shall take up the various designs involved in experimental method.

5.4.3 Designs of Experimental Study

A research design is very important for the researcher. A well developed design provides the structure and strategy to control the investigation and extract dependable answers to the questions raised by the problem or, hypothesis. It is the nature of the problem that determines the appropriateness of the design.

Before we discuss the experimental designs, it will be relevant to look into the terms and symbols which we shall make use of.

i) X represents the *independent variable,* which is manipulated by the researcher, it is also referred to as the *experimental variable* or the *treatment variable.*

ii) Y represents the measure of the *dependent variable.* Y1 represents the dependent variable before the manipulation of the independent variable X; it is usually a pre-test of some type administered before the experimental treatment. Y2 represents the dependent variable after the manipulation of the independent variable X; it is usually a post-test administered to subjects after the experimental treatment.

iii) S represents the *subject* or *respondent* used in the experiment.

iv) E group refers to the *experimental group* – the group that is given the independent variable treatment.

v) C group refers to the *control group* – the group that does not receive the experimental treatment.

vi) R indicates *random assignment* of subjects to the experimental groups and the random assignment of treatments to the groups.

There is a large number of experimental designs. Various authors have classified experimental designs into certain categories. Most common categorization comprises:

- Pre-experimental Design
- True Experimental Design
- Quasi Experimental Design

Some authors like Donald Ary and others (1985) add two more categories namely

- Factorial Design
- Time Series

Various designs under the above mentioned categories are given in the Table 5.1

Table 5.1: Various experimental designs

Pre-experimental	True Experimental	Quasi Experimental	Factorial Design	Time Series
• One Group pre-test Post-test Design • Static Group comparison	• Randomized Subjects' Post-test only Control Group Design • Randomized Matched Subjects, Post-test only Control Group Design • Randomized subjects pre-test Post-test control Group Design • Solomon Three Group Design • Solomon Four Group design	• Nonrandomized Control Group, pre-test Post-test Design • Counter Balanced Design	• Simple Factorial Design	• One Group Time Series Design • Control Group Time Series Design

However, in this section, we will bring before you only a few most frequently used designs, from each of the five categories. We begin with the pre-experimental design.

A. *Pre-experimental Design*

The two designs classified as pre-experimental designs offer minimal control of extraneous variables. Still they are used quite often for research in the area of nutrition health. These designs help to illustrate the advantages of more rigorously controlled designs that are presented later.

Design 1: One Group Pre-test Post-test Design

When this design is employed, the dependent variable is measured before the independent variable or treatment is applied or withdrawn, an then measured yet again. The one group design usually involves three steps:

a) administering a pre-test measuring the dependent variable,

b) applying the experimental treatment X to the subject, and

c) administering a post-test again measuring the dependent variable.

Differences attributed to application of experimental treatment are then determined by comparing the pre-test and post-test scores.

Pre-test	Independent Variable	Post-test
Y_1	X	Y_2

Design 1: One Group Pre-test Post-test Design

To illustrate the use of this design, let us as researchers assume that we want to evaluate the effectiveness of a particular nutrition education programme to improve iron status. How may we go about this task?

At the beginning of the research, the subjects are given a standardized test that measures the knowledge of the subjects regarding say anaemia quite satisfactorily, following which the researcher then introduces the nutrition education material. At the end of the study, the subjects are administered the standardized test a second time. Comparing the scores of the two tests would reveal what difference the exposure to the nutrition education has made.

However, using only one group, as in Design 1, gives us superficial control. The major limitation of the one-group design is that, since no control group is used, the experimenter cannot assume that the change between the pre-test and the post-test scores is brought about by the experimental treatment alone. It is quite possible that some extraneous variables account for all or part of the change. For example, subjects experience changes with the passage of time; they grow mentally, as well as, physically, or they may acquire additional learning experiences that would affect the dependent variable. This extraneous variable can be though of as *maturation* i.e. with the passage of time subjects get maturity and this in turn may effect achievement level. Another type of extraneous variable that can operate between the pre-test and the post-test scores and which cannot be controlled is *history*. History as a source of extraneous variance refers to the specific events that can occur between the pre-test scores and the post-test other than the experimental treatment. In the example cited above, not receiving the nutrition education regularly or illness just before the test could decrease achievement scores. Similarly, a crucial discovery regarding anaemia could increase widespread interest and hence affect that test scores. In fact, history and maturation become increasingly influential sources of extraneous variance when the time interval between Y1 and Y2 is long.

Another short coming of Design 1 is that it offers no way of assessing the effect of the pre-test Y1 itself. We know that "practice effect" exists when subjects take a test a second time or take an alternate form of the test. In other words, subjects do better the second time even without any instruction or relevant discussion during the interval. This is true not only for achievement and intelligence tests but also for personality tests.

To sum up, Design 1 has little to recommend it; without a control group to make a comparison possible, the results obtained in a one group design are basically

uninterruptible. The results of the experiment would have been dependable if there could be a comparable group i.e. control group to which nutrition education had not been given.

Design 2: Static Group Comparison

Design 2 utilizes two or more groups, only one of which is exposed to experimental treatment. The groups are assumed to be equivalent in all relevant aspects, they differ only in their exposure to X.

This design is often used in nutrition education research. For example, the achievement of subjects taught by a new method (multi-media) is compared with that of a similar class taught by a traditional method (lecture).

Design 2 has a control group or groups, which permit(s) the comparison that is required for scientific respectability. If the experimental group is superior on the Y2 measure, the researcher then has more confidence in his/her conclusion that the difference is due to experimental treatment.

However, there is a basic flaw in this design. Since neither *randomization* (refer to Unit 6 for details) nor even *matching* is used to assign subjects to the experimental and control groups, we cannot be sure that the groups are equivalent prior to the experimental treatment. They may differ on certain relevant variables, and it may be these differences rather than X that are responsible for the observed change. Because we cannot be sure that the groups are equal with regard to all the factors that may influence the dependent variable, this design is considered to be lacking in the necessary control and must be classified as pre-experimental.

Group	Independent Variable	Post-test
E	X	Y2
C	–	Y2

Design 2: Static Group Comparison

Next, let us get to know about the true experimental designs.

B. ***True Experimental Designs***

The following two design belong to the 'true experimental' design, because of the control that they provide. i.e.

i) Random assignment of subjects to the groups

ii) Random assignment of treatment in the groups

iii) Post-testing all the groups.

Let us consider Design 3 presented herewith.

Design 3: Randomized Subjects, Post-test only Control Group Design

This particular design requires two groups to which subjects are randomly assigned and each group is assigned to a different condition. No pre-test is used; randomization controls all the possible extraneous variables. This does not mean that randomization procedures (like drawing names out of a hat, or flipping a coin) remove the extraneous variables, such as the IQ or age, which may affect the dependent variable, or control their presence. These extraneous variable still affect the inquiry, but now it is the laws of chance rather than the personal features of E that operate. In fact, the larger the number of subjects used, the more equivalent or similar the groups will tend to be. Suppose a researcher wants to study the effect of a supplement on disease condition in a research study. He/she may randomly assign the subjects to the groups and

provide treatment to one of the groups. The assigning of the treatment will be random. At the end of the study he/she may test both the groups.

After the subjects are assigned to the groups, only the experimental group is exposed to the experimental treatment (i.e. the supplement). Otherwise, in all other respects, the two groups remain similar. Members of both groups are then measured on dependent variable Y2. Scores are then compared to determine the effect of X.

Group	Independent Variable	Post-test
(R)E	X	Y2
(R)C	–	Y2

Design 3: Randomized Subjects, Post-test only Control Group Design

The main advantage of Design 3 is randomization, which assures statistical equivalence of the groups prior to the introduction of independent variable Design 3 provides controls for the main effects of history, maturation and pretesting because no pre-test is used, there can be no interactional effect of pre-test and X (treatment).

Next, let us consider Design 4.

Design 4: Randomized Matched Subjects, Post-test only Control Group Design

This design is similar to Design 3 except that it uses a matching techniques, rather than random assignment to obtain equivalent groups. Subjects are matched on one or more variable that can be measured conveniently, such as age, socio-economic, educational status etc. The matching variables used are generally those that have a significant correlation with the dependent variable. On the basis of these variables, subjects are paired so that opposite members/scores are as close as possible; and then, one member of each pair is randomly assigned to one treatment and the other to the second treatment.

Group	Independent Variable	Post-test
(Mr) E	X	Y_2
(Mr) C	–	Y_2

Design 4: Randomized Matched Subjects, Post-test only Control Group Design

Matching is most useful in studies where small samples are to be used and where Design 3 is not appropriate. Also, the matched subjects' design serves to reduce the extent to which experimental differences can be accounted for by initial differences between groups. However, for matching to really become a means of control, the matching of all the potential subjects must be complete, and the assignment of the members of each pair to the groups must be determined randomly. If one or more subjects should be excluded because an appropriate match could not be found, this would bias the sample. When using Design 4, it is essential to match every subject, even if only approximately, before random assignment is effected.

Next, let us review the quasi experimental design.

C. *Quasi Experimental Design*

One of the quasi experimental designs is Non-randomized Control Group, Pre-test Post-test Design. You would notice that randomized control group pre-test post-test design is a true experimental design which we have presented before. The only difference in the quasi experimental design is that the groups are not randomized. Hence, they are unlikely to be comparable. In fact, it is on this ground that the design becomes quasi experimental and not true experimental. Since the rest of the design related characteristics remain common with the randomized control group pre-test

post-test design of the true experimental design category, we do not need to provide any further details on this design.

Group	Pretest	Independent Variable	Post-test
E	Y_1	X	Y_2
C	Y_1	–	Y_2

Finally let us review the factorial and the time series design.

D. ***Factorial Designs***

A factorial design is one where two or more variables are manipulated simultaneously in order to study the independent effect of each variable on the dependent variable, as well as, the effects due to interactions among the several variables. Factorial designs are of two types. In the first type, one of the independent variables may be experimentally manipulated. The researcher is primarily interested in the effect of a single independent variable but he/she must take other variables into consideration which may influence the dependent variables. In the second type of design, all the independent variables may be experimentally manipulated. Factorial designs have been developed at varying levels of complexity. The simplest factorial design is the 2 by (2 × 2) Design. The two independent variables have two values.

Level 1 subjects receive Treatment A and others Treatment B. Some level 2 subjects receive Treatment A and other Treatment B.

Attribute Variable X_2	Experimental Variable X_1	
	Treatment A	Treatment B
Level 1	Cell 1	Cell 3
Level 2	Cell 2	Cell 4

The strength of the factorial design is that it can achieve in one experiment what might otherwise require two or more separate studies.

E. ***Time Series Design***

We have already discussed pre-test post-test designs. They generate one time data on the dependent variable before and after the experimental treatment. There are instances in epidemiological research where it becomes necessary to compare change in the trend of a particular phenomenon or process or product. For example, let us assume that subjects behaviour attitudes, achievements etc. changes over a period of time. If a specific treatment is introduced in a research protocal to study the change in attitude or achievement it is useful to study the trend through measurement at certain intervals before the introduction of the treatment. Instead of one time pre-test, the test is repeated three or four times before the treatment is administered. This generates data on the trend of behaviour. Similarly after the treatment is administrered instead of one time post-test, the post-test is administered several time at intervals. This provides data to derive the trend in the change in behaviour. Since both, pre-tests and post-tests are used over a time, it is called *Time Series Design*. In a time series design the effect of the treatment on the dependent variable is tested by comparing the trends. This can be represented in the following:

$$Y_1 \quad Y_2 \quad Y_3 \quad Y_4 \quad X \quad Y_5 \quad Y_6 \quad Y_7 \quad Y_8 .$$

What we have described above is one group time series design. If you add a control group and repeat the same time series measurement without the treatment of the control group it becomes control group time series design. Similarly control group time series design is represented as.

Group									
E	Y_1	Y_2	Y_3	Y_4	X	Y_5	Y_6	Y_7	Y_8
C	Y_1	Y_2	Y_3	Y_4	-	Y_5	Y_6	Y_7	Y_8

Check Your Progress Exercise 2

1) Explain briefly the significance of control, manipulation and observation in an experimental study.

.....

.....

.....

2) Draw and compare the figures representing pre-test post-test experimental design and one group time series design.

.....

.....

.....

.....

.....

With a brief review of time series design we end our study of the designs of experimental studies. Now, surely you will be in a better position to select and conduct experimental designs to suit your research.

5.5 LET US SUM UP

The main focus of this unit was to orient the learners to the basic characteristics and features of descriptive and experimental research and elaborate on the steps involved in the designing and conduct of these epidemiological research designs. The various designs of experimental studies was explained giving appropriate examples.

5.6 GLOSSARY

Post-test : a test designed to measure the individuals level of understanding of a concept/process after some internvention (education etc).

Pre-test : a test designed to measure the subjects general level of understanding for some basic concept/ process term. It is a test taken before the study intervention.

Quasi Experimental Design : an experimental design that deos not meet all the requirements necessary for controlling the influence of extraneous valuables. They do not have randomly assigned treatment and comparison groups. Instead the comparison between treatment and nontreatment conditions is made with nonequivalent groups or with the same subjects prior to treatment.

5.7 ANSWERS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress Exercise 1

1) Correlational studies are useful to:

   - determine the relationship between variables and
   - measure the extent to which variations in one variable are associated with the variations in another variable. The weaknesses of causal comparative studies are: lack of control, difficulty in identifying the relevant causal factors, determining their number in given phenomenon, classifying subjects into dichotomous groups for the purpose of comparison, lack of control over the selection of subjects.

2) Refer to sub-section 5.2.3 and answer on your own understanding of the topic.

Check Your Progress Exercise 2

1) Control is crucial to (i) evaluate unambiguously the effects of an independent variable and (ii) arrange a situation in which the effect of variables can be measured.

   Manipulation controls or eliminates the irrelevant factors and arranges a situation in which only relevant factors can be studied.

   Observations are made to study specific characteristics in the behaviour of the subjects employed in experimental research.

2) Pre-test Post-test Experimental Design

Group	Pre-test	Treatment	Post-test
E	Y_1	X	Y_2

One Group Time Series Design

Y_1	Y_2	Y_3	Y_4	Y	Y_5	Y_6	Y_7	Y_8

Compared to one test each before and after the treatment in pre-test post-test experimental group design, tests are repeated at specified intervals in one group time series design. Whereas time series designs compares the trends of change in the dependent variable, the pre-test post-test experimental design tests one time gain or change in the dependent variable.