**CHAPTER FIVE: RESEARCH METHODS**

**5.1 Learning objectives**

**After completing this chapter, the student should be able to**:

1. Identify the pertinent questions to consider when developing the methodology of a research proposal
2. Describe and understand the various components of the methods section in a research proposal
3. Explain the cyclical nature of the different steps in designing the methodology.

**5.2 Introduction**

In the previous chapters we have dealt with the identification, selection, analysis and statement of the problem. The importance of literature review and formulation of study objectives were also emphasized. Now we must decide exactly how we are going to achieve our stated objectives. That is, what new data do we need to shed light on the problem we have selected and how we are going to collect and process these data? The major issues that constitute the "methods section" of a research proposal will be dealt in the sections that follow.

**5.3 Types of study designs**

A study design is the process that guides researchers on how to collect, analyze and interpret observations. It is a **l*ogical model*** that guides the investigator in the various stages of the research.

Several classifications of study types are possible, depending on what research strategies are used.

1. **Non-intervention (Observational) studies** in which the researcher just observes and analyses researchable objects or situations but does not intervene; and
2. **Intervention studies** in which the researcher manipulates objects or situations and measures the outcome of his manipulations (e.g., by implementing intensive health education and measuring the improvement in immunization rates.)

**Study designs could be exploratory, descriptive or analytical**

**1. Exploratory studies**

An **exploratory study** is a small-scale study of relatively short duration, which is carried out when little is known about a situation or a problem. It may include description as well as comparison.

**For example:**

A national AIDS Control Program wishes to establish counseling services for

HIV positive and AIDS patients, but lacks information on specific needs patients have for support. To explore these needs, a number of in-depth interviews are held with various categories of patients (males, females, married and single) and with some counselors working on a program that is already under way.

When doing exploratory studies we *describe* the needs of various categories of patients and the possibilities for action. We may want to go further and try to explain the differences we observe (e.g., in the needs of male and female AIDS patients) or to identify causes of problems. Then we will need to *compare* groups.

**If the problem and its contributing factors are not well defined it is always advisable to do an exploratory study before embarking on a large-scale descriptive or comparative study.**

**2. Descriptive studies:**

Descriptive studies may be defined as studies that describe the patterns of disease occurrence and other health-related conditions by ***person place*** and ***time.***

**Personal variables include**: basic demographic factors, such as age, sex marital status or occupation, as well as the consumption of various types of food or medication use.

Characteristics of place refer to the **geographic distribution of disease*,*** including variation among countries or within countries, such as between urban and rural areas.

With regard to time, descriptive studies may examine **seasonal patterns in disease onset, etc.**

**Uses of descriptive studies**

* They can be done fairly quickly and easily.
* Allow planners and administrators to allocate resources
* Provide the first important clues about possible determinants of a disease (useful for the formulation of hypotheses)

**Types of descriptive studies**

***a) Case reports and case series***

**Case report:** a careful, detailed report by one or more clinicians of the ***profile of a single patient.***

The individual case report can be expanded to a ***case series,*** which describes characteristics of a number of patients with a given disease.

Uses

* Important link between clinical medicine and epidemiology
* One of the first steps in outbreak investigation
* Often useful for hypothesis generating and examining new diseases, but conclusions about etiology cannot be made.

**b) Ecological studies:** data from entire populations are used to compare disease frequencies between different groups during the same period of time or in the same population at different points in time.

**Example:** Countries with low cigarette consumption have lower lung cancer rates than those countries with high cigarette consumption.

* Ecological studies are usually quick and easy to do and can be done with already available information.
* Since ecological studies refer to whole populations rather than to individuals, it is not possible to link an exposure to occurrence of disease in the same person.

**c) Cross-sectional studies**

A cross-sectional (prevalence) study provides information concerning the situation at a given time. In this type of study, the status of an individual with respect to the presence or absence of both exposure and disease is assessed at the same point in time.

* Usually involve collection of new data.
* In general, measure prevalence rather than incidence
* Not good for studying rare diseases or diseases with short duration; also not ideal for studying rare exposures.

For factors that remain unaltered over time, such as sex, blood group, etc., the cross sectional survey can provide evidence of a valid statistical association.

As can be noted from the above explanation, a cross-sectional study can be either analytical or descriptive, according to its purpose. If data are collected both on exposures and outcomes of interest, and if the data are analyzed so as to demonstrate differences either between exposed and non-exposed groups, with respect to the outcome, or between those with the outcome and those without the outcome, with respect to the exposure, then this is an analytical cross-sectional study. If the information collected is purely of a descriptive nature, not involving the **comparison** of groups formed on the basis of exposure or outcome status, then this is a descriptive cross-sectional study. Often a cross-sectional study may have both descriptive and analytical components.

Nowadays, there is an increasing emphasis on the value of longitudinal studies in which observations are repeated in the same community over a prolonged period (i.e., longitudinal studies provide the required data at **more than one point** in time unlike cross- sectional surveys)**.**

**II. Analytic studies**

Analytic studies may be defined as studies used ***to test hypotheses*** concerning the relationship between a suspected risk factor and an ***outcome*** and to **measure the magnitude of the association** and its ***statistical significance.***

Analytic study designs can be divided into two broad design strategies: Observational and intervention.

**Observational studies**

* No human intervention involved in assigning study groups; simply observe the relationship between exposure and disease.
* Subject to many potential biases, but by careful design and analysis, many of these biases can be minimized.
* Examples of observational studies: comparative cross-sectional, cohort and case control studies.

**a) Comparative cross-sectional studies:** Depending on the purpose of a given study, a cross-sectional survey could have an analytical component (see section 6.3, 2c, above).

**b) Cohort studies:** Study groups identified by exposure status prior to ascertainment of their disease status and both exposed and unexposed groups followed in identical manner until they develop the disease under study, they die, the study ends, or they are lost to follow-up.

**Strengths and limitations of the cohort study design**

**Strengths:**

* Is of particular value when the **exposure is rare**
* Can examine **multiple effects** of a single exposure
* Allows direct measurement of **incidence** of disease in the exposed and non-exposed groups.

**Limitations:**

* Is inefficient for the evaluation of rare diseases
* Expensive and time consuming
* Validity of the results can be seriously affected by losses to follow-up.

**c) Case-control studies:** Group of subjects with the disease (cases) and group of subjects without the disease (controls) are identified. Information, about previous exposures is obtained for cases and controls, and frequency of exposure compared for the two groups.

**Strengths and limitations of the case-control study design**

**Strengths:**

􀂃 is relatively quick and inexpensive

􀂃 is optimal for the evaluation of rare diseases.

􀂃 can examine multiple etiologic factors for a single disease.

**Limitations:**

* Is inefficient for the evaluation of rare exposures
* Cannot directly compute incidence rates of disease in exposed and non- exposed individuals.
* Is particularly prone to bias compared with other analytic designs, in particular, selection and recall bias.

**Intervention studies**

In intervention studies, the researcher manipulates a situation and measures the effects of this manipulation. Usually (but not always) two groups are compared, one group in which the intervention takes place (e.g. treatment with a certain drug) and another group that remains ‘untouched’ (e.g. treatment with a placebo).

The two categories of intervention studies are:

* experimental studies and
* quasi-experimental studies

**1. Experimental studies**

An experimental design is a study design that gives the most reliable **proof for causation**. In an **experimental study,** individuals are randomly allocated to at least two groups. One group is subject to an intervention, or experiment, while the other group(s) is not. The outcome of the intervention (effect of the intervention on the dependent variable/problem) is obtained by comparing the two groups. A number of experimental study designs have been developed.

These are widely used in laboratory settings and in clinical settings. For ethical reasons, the opportunities for experiments involving human subjects are restricted. However, randomized control trials of new drugs are common.

At community level, where health research is frequently undertaken, we experience not only ethical but also practical problems in carrying out experimental studies. In real life settings, it is often impossible to assign persons at random to two groups, or to maintain a control group.

Therefore, experimental research designs may have to be replaced by quasi-experimental designs.

**2. Quasi-experimental studies**

In a **quasi-experimental study**, one characteristic of a true experiment is missing, either randomization or the use of a separate control group. A quasi-experimental study, however, always includes the manipulation of an independent variable which is the intervention.

One of the most common quasi-experimental designs uses two (or more) groups, one of which serves as a control group in which no intervention takes place. Both groups are observed before as well as after the intervention, to test if the intervention has made any difference. (This quasi-experimental design is called the ‘non-equivalent control group design’ because the subjects in the two groups (study and control groups) have not been randomly assigned.)

Another type of design that is often chosen because it is quite easy to set up uses only **one group** in which an intervention is carried out. The situation is analyzed before and after the intervention to test if there is any difference in the observed problem. This is called a ‘BEFORE-AFTER’ study. This design is considered a ‘pre-experimental’ design rather than a‘quasi-experimental’ design because it involves neither randomization nor the use of a control group.

**Intervention (experimental) studies can also be considered either therapeutic or Preventive**.

Therapeutic trials are conducted **among patients with a particular disease** to determine the ability of an agent or procedure to diminish symptoms, prevent recurrence, or decrease risk of death from that disease.

A preventive trial (community trial) involves the evaluation of whether an agent or ***procedure reduces the risk of developing disease*** among those ***free from that*** condition at enrolment. Thus, preventive trials can be conducted among individuals at usual risk (e.g. Vaccine trials)

A particular research question may be addressed using different approaches. The choice of study design for investigation is influenced by:

* Particular features of the exposure and disease.
* Logistic considerations of available resources.
* Results from previous studies and gaps in knowledge that remain to be filled.
* Ingenuity and creativity of the researcher

**6.4 Study population**

At an early stage in the planning of any investigation decisions must be made concerning the study population. That is, concerning the population of individual units (whether they are persons, households, etc.) to be investigated. The population under consideration should be clearly and explicitly defined in terms of place, time, and other relevant criteria. If the study population comprises cases of a disease the procedures to be used for case identification should be stated. If controls are to be chosen their method of selection should be stated.

Often the investigator will have implicitly chosen his study population when he defined the topic of his investigation, by reason of his interest in a specific community or a specific health program.

In other instances, particularly when an analytic survey or an experiment is being planned, the investigator may require purposively to select a study population. In so doing he must consider questions of appropriateness and practicability.

The appropriateness of the study population refers to its suitability for the attainment of the objectives of the study.

The selection of study population on the basis of suitability usually affects the validity of subsequent generalizations from the findings. This situation requires a close attention at the early stage of the given study. Two examples are given below.

a) **Volunteer populations**: Persons who volunteer to enter a study may differ in many respects from those who do not so volunteer, and therefore the findings in a volunteer population do not necessarily apply to the population at large.

b) **Hospital or clinic populations**: Persons receiving medical care are obviously not representative of the general population from which they have come from. That is, persons treated in hospital for a certain disease may differ from those patients with the same disease but not receiving care for it.

**Practical questions such as the following could also arise**.

- Is the proposed population the one that would give the required information?

- Will the population cooperate to participate in the study, or will it be a 'resistant' one?

- If it is proposed to study patients with a specific disease, will it be possible to identify enough cases to yield useful conclusions?

- If a long term 'follow up' study is planned, is the population so mobile that it may be difficult to maintain contact with the subjects?

**A preliminary exploratory study may sometimes be required in order to answer such questions.**

**5.5 Operational Definitions of Variables**

Before we directly go to the operational definition of variables it would be important to discuss about the nature of variables first.

**Definition**: A variable is a characteristic of a person, object, or phenomenon that can take on different values.

A simple example of a variable is a person's age. The variable can take on different values, such as, 20 years old, 30 years old, and so on. Other examples of variables are:

a) Weight in kilograms

b) Height in centimeters

c) Monthly income in Birr

d) Marital status (single, married, divorced and widowed)

e) Job satisfaction index (1 to 5)

f) Occupation (civil servant, farmer, student, ET.)

g) Disease condition (presence or absence of a disease)

The first three variables (a to c) are **numerical** variables because they are expressed in numbers (metric data). Since the values of the remaining three variables (d to g) are expressed in categories, we call them **categorical** variables.

Because in the health research we often look for associations, it is important to make a distinction between **dependent and independent** variables. Both the dependent and independent variables together with their operational definitions (when necessary) should be stated.

**Definitions:**

The variable that is used to describe or measure the problem under study is called the dependent variable. The variables that are used to describe or measure the factors that are assumed to influence (or cause) the problem is called independent variables.

For example, in a study of relationship between smoking and lung cancer, "suffering from lung cancer" (with the values yes, no) would be the **dependent** variable and "smoking" (with the values no, less than a packet/day, 1 to 2 packets/day, more than 2 packets/day) would be the **independent** variable.

**Background variables** - In almost every study involving human subjects, background

Variables, such as, age, sex, educational status, monthly family income, marital status and religion will be included. These background variables are often related to a number of independent variables, so that they influence the problem indirectly. Hence they are called background variables or background characteristics.

**Confounding variable** - A variable that is associated with the problem and with a possible cause of the problem is a potential confounding variable. This type of variable may either strengthen or weaken the apparent relationship between the problem and a possible cause.

**Composite variable** - A variable based on two or more other variables may be termed a composite variable. Incidence and prevalence rates, sex ratios, and other rates and ratios are composite variables, since they are based on separate numerator and denominator information.

**I. Operationalizing variables by choosing appropriate indicators**

Note that the different values of many of the variables presented above can easily be determined. However, for some variables it is sometimes not possible to find meaningful

Categories unless the variables are made operational with one or more precise

**INDICATORS** - Operationalising variables means that you make them ‘measurable'.

**For example:**

1. In a study on VCT acceptance, you want to determine the **level of knowledge** concerning HIV in order to find out to what extent the factor ‘poor knowledge’ influences willingness to be tested for HIV. The variable ‘level of knowledge’ cannot be measured as such. You would need to develop a series of questions to assess a person’s knowledge, for example on modes of transmission of HIV and its prevention methods. The answers to these questions form an **indicator** of someone’s knowledge on this issue, which can then be categorized. If 10 questions were asked, you might decide that the knowledge of those with:

— 0 to 3 correct answers is poor,

— 4 to 6 correct answers is reasonable, and

— 7 to 10 correct answers are good.

When defining variables on the basis of the problem analysis diagram, it is important to realize which variables are measurable as such and which ones need indicators. Once appropriate indicators have been identified we know exactly what information we are looking for. This makes the collection of data as well as the analysis more focused and efficient.

1. Nutritional status of under-5 year olds is another example of a variable that cannot be measured directly and for which you would need to choose appropriate indicators. Widely used indicators for nutritional status include weight for age, weight for height, and height for age, and upper-arm circumference. For the classification of nutritional status, internationally accepted categories already exist, which are based on standard growth curves. For the indicator weight/age, for example, children are:
* Well-nourished if they are above 80% of the standard
* Moderately malnourished if they are between 60% and 80%
* Severely malnourished if they are below 60%

**II. Defining variables and indicators of variables**

To ensure that everyone (the researcher, data collectors, and eventually the reader of the research report) understands exactly what has been **measured** and to ensure that there will be consistency in the **measurement,** it is necessary to clearly define the variables (and indicators of variables). For example, to define the indicator “waiting time” it is necessary to decide what will be considered the starting point of the “waiting period” e.g. Is it when the patient enters the front door, or when he has been registered and obtained his card?

For certain variables, it may not be possible to adequately define the variable or the indicator immediately because further information may be needed for this purpose. The researcher may need to review the literature to find out what definitions have been used by other researchers, so that he can standardize his definitions and thus be able later to easily compare his findings with those of the other studies. In some cases the opinions of “experts “or of community members of health care providers may be needed in order to define the variable or indicator.

The variables to be studied are selected on the basis of their relevance to the objectives of the investigation.

* **the initial list is usually too long**
* **it has to be pruned to facilitate the collection and processing of the data.**

Once the variables are selected, each of them should be clarified. There are two aspects to be considered.

1. Clear definition of variables in terms of **objectively measurable facts** (i.e., operational definition) - this was repeatedly mentioned (addressed) in the above examples

2. The **scale of measurement** to be used in data collection.

Unless the variables are clearly and explicitly defined, there can be no assurance that, if the study is performed by a different investigator, or repeated by the same investigator, similar findings would be obtained.

The following example shows the different definitions (two different definitions) given to "obesity".

The two kinds of definitions are: conceptual and operational. The conceptual definition is often akin to a dictionary definition. **e.g.** “Obesity” may be defined as: “excessive fatness”, “overweight”, etc.

In contrast, the operational definition is heavily influenced by considerations of practicability. “Obesity”, for example, might be operationally defined as: “ a weight, based on weighing in underclothes and without shoes, which exceeds, by 10% or more, the mean weight of persons of the subject's sex, age and height (in a specified population at a specified time)".

In general, operational definitions of variables are used in order to:

􀂃**Avoid ambiguity**

􀂃**Make the variables to be more measurable**