# UNIT 1 BASIC CONCEPTS

**Structure**



## 1.1 INTRODUCTION

This first unit in the course will orient you to the basic concepts, constructs and principles in scientific approach to research. Starting with a basic understanding of epidemiology, research and biostatistics, the unit will focus on the importance of scientific approach in research and discuss the nature of research in epidemiology especially with reference to nutrition and medicine.

**Objectives**

After studying this unit, you will be able to:

- define the terms biostatistics, epidemiology, research,
- explain the importance of scientific approach,
- name and describe characteristics of the epidemiologic approach,
- define the components of epidemiology,
- discuss the nature of research in epidemiology especially with reference to nutrition and medicine,
- explain giving examples what is meant by natural experiments, and
- enumerate the areas in which nutrition/health research is undertaken.

## 1.2 EPIDEMIOLOGY: AN INTRODUCTION

Hardly a day passes without some study being publicized in newspapers about foods, new infections, food additives, medicines, drugs, chemical wastes or pollutants. However, it is difficult to know whether these reports are genuine or are based on half-baked and preliminary findings. Some studies give us a lot of hope whereas others give us anxiety. Many contradict each other resulting is confusion. Much of this can be avoided if we have understanding of the power and limitations of scientific studies especially epidemiology.

Let us first get some basic concepts clear which most researchers are required to understand and use.

Research especially in nutrition and medicine requires ability to design and conduct the study, biological understanding, statistical expertise and many other skills that vary from one study to another. Our understanding of biologic mechanism is far too incomplete to predict confidently the ultimatê consequences of eating a particular food or nutrient. A good research essentially is one where the research study is conducted and described in such a way that anyone who is interested can follow the argument and decide for oneself the validity of the conclusions. Research studies help us to describe, explain, predict and control.

Epidemiologic studies directly relating intake of dietary factors or nutrients to risk of death or disease among humans play a critical complementary role to laboratory investigations. You must have noticed that in the last 3-4 decades, there have been numerous reports in the literature on nutritional epidemiology. This has led to the development of a firmer quantitative basis for this science. What is nutritional epidemiology? We shall understand the concept in the next section.

### 1.2.1 Epidemiology: Concept and Definitions

Let us begin by first defining the term epidemiology. Epidemiology is derived from the Greek words: *Epi* (meaning upon, among) + *demos* (people, population) + *logos* (study of, discourse). Epidemiology is the simplest and most direct method of studying the cause of disease in humans. It is a discipline that describes, quantifies and postulates causal mechanisms for health phenomena. A comprehensive definition of epidemiology would be *'the study of the distribution and determinants of disease frequency'* in human population.

Let us look at this definition more closely. You would have noticed that epidemiology concerns with the three components namely – *distribution, determinants and disease frequency.* Let us review these components.

*Determinants* are factors or events that are capable of bring about a change in health. Examples of determinants would include say bacteria like mycobacterium tuberculosis, which causes TB, the AIDS virus, chemical agents that may be carcinogens (cancer–causing). Others may be less specific e.g. lack of fat, diets high in saturated fats etc. *Determinants* are all the physical, biological, behavioural, social, and cultural factors that influence health.

*Distribution*: Distribution involves analysis of data according to the time scale over which events occur, the places where the events occur, and the categories of persons to whom they occur. The distribution of disease considers such questions as who is getting the disease within a population group, as well as, where and when the disease is occurring. The disease distribution may vary from one population group to another e.g. diabetes may vary from rural population group to urban upper income group. Degenerative diseases like obesity may be more prevalent in an urban setting, linked to sedentary lifestyles as compared to rural background characterized by active life stylès.

*Frequency*: The frequency of disease occurrence or the measurement of disease frequency, involves quantification of the existence or occurrence of disease.

In 1983 a committee representing the *International Epidemiological Association* defined epidemiology as *"the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems."*

From our discussion above, it must be evident that *epidemiology* is concerned with the frequencies and types of illnesses/diseases and injuries in groups of people and with the factors that influence their distribution. Put in simple words, it deals with the study of the causes, distribution, and control of disease in populations.

Epidemiology is based on two fundamental assumptions. These are:

1) Human disease does not occur at random.

2) Human disease has causal and preventive factors that can be identified through systematic investigation of different populations or subgroups of individuals within a population in different places or at different times.

We can go beyond health and diseases – can study disability, morbidity and mortality. *Morbidity*, you may already be aware, designates illness. *Mortality* refers to deaths that occur in a population or other group. *Disability* refers to a state of being disabled (mentally, physically). Most measures of morbidity and mortality are defined for specific types of morbidity or causes of death. Thus one definition of epidemiology *is the study of the nature, cause, control, and determinants of the frequency and distribution of disease, disability, and death in human populations.*

The interrelatedness of the various factors that can contribute in the investigation of diseases and epidemics can be better understood in the form of the triangle of epidemiology. Refer to Figure 1.1 which illustrates the triangle of epidemiology.

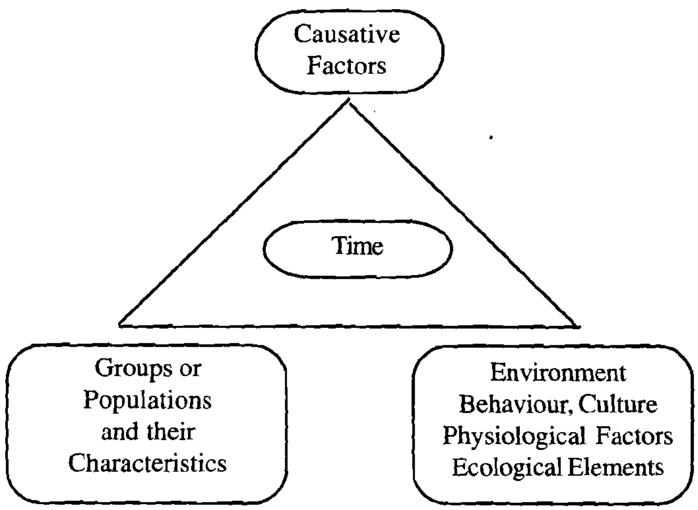


**Figure 1.1: The triangle of epidemiology**

Figure 1.1 illustrates the multifaceted phenomenon in the epidemiology of disease. It shows the interaction and interdependence of causative factors, characteristics of population groups and environment, behaviour, lifestyle issues and time as used in the investigation of diseases and epidemics. The model presented above recognizes that disease states and conditions affecting a population are complex and that causative factors are many. It gives recognition to the fact that many factors and elements contribute to disease and maladies in populations, better reflecting the behaviour, lifestyle, and chronic disease issues found in modern times. The concept of *causative factors* implies the need to identify multiple causes or etiologic factors of disease, disability, injury and death.

With a basic understanding of epidemiology now let us consider nutritional epidemiology. The field of *nutritional epidemiology* has developed specifically from interest in the concept that aspect of diet may influence the occurrence of human disease.

*Nutritional epidemiology* deals with the study of the relationship, if any, between nutrient/dietary intake and disease condition. To illustrate, study of relationship, if any, between dietary fat intake and breast cancer incidence. It is important to understand here that the effects of dietary intake and nutritional status on health are complex. Multiple nutrients work together in the body, with differing impacts on different systems and under differing environmental conditions. The effects of dietary intake may be modified or confounded by other exposures, including physical activity, smoking, and alcohol consumption. Understanding and untangling specific effects requires an understanding of the complex interactions among exposures. Nutritional epidemiology deals with understanding this complex phenomenon.

Next, let us review the aim/goals of epidemiology.

### 1.2.2 Epidemiology: Purpose, Aim and Goals

To describe health/nutrition status of populations means to enumerate the cases of disease, to obtain relative frequencies of a disease within subgroups, to discover important trends in occurrence of the disease. The purpose of epidemiology, therefore, is:

- to *explain etiology* i.e. to discover causal factors or modes of transmission,
- to *predict occurrence* – means to estimate the actual number of cases that will develop, as well as, to identify the distribution within populations. This information is crucial to planning interventions and allocation of resources, and
- to *control* – to prevent occurrence of new cases, to eradicate existing cases and to prolong the lives of those with the disease.

We may elaborate further by saying that the aims of epidemiology are to:

- determine the primary agent or ascertain causative factors,
- understand the causation of disease, disorders, or conditions,
- determine the characteristics of the agent or causative factors,
- define the mode of transmission,
- define and determine contributing factors,
- identify and explain geographic disease patterns,
- determine, describe and report the natural course of disease, disability, injury and death,
- determine control methods,
- determine preventive measures,
- aid in the planning and development of health services, and
- provide administrative and planning data.

Several uses of epidemiology have been described. These include:

1) To *study the history of the health of populations*, and of the rise and fall of diseases and changes in their character. Useful projections into the future may be possible.

2) To *diagnose the health of the community* and the conditions of the people, to measure the true dimensions and distribution of ill-health in terms of incidence, prevalence, disability and mortality; to set health problems in perspective and define their relative importance; to identify groups needing special attention. Ways of life change, and with them the community's health; new measurements for monitoring them must therefore constantly be sought.

3) To *study the working of health services* with a view to their improvement. Operational research translates knowledge of (changing) community health and expectations in terms of needs for services and measure how these are met. The

success of services delivered in reaching stated norms, and the effects of community health and its needs have to be appraised, in relation to resources. Such knowledge may be applied in action research pioneering better services, and in drawing up plans for the future. Timely information on health and health services is itself a key service requiring much study and experiment. Today, information is required at many levels, from the local district to the international.

4) To *estimate* from the group experience what are *the individual risks* on average of disease, accident and defect, and the chances of avoiding them.

5) To *identify syndromes* by describing the distribution and association of clinical phenomena in the population.

6) To complete the *clinical picture of chronic diseases* and describe their natural history: by including in due proportion all kinds of patients, wherever they are present, together with the undemanding and the symptom-less cases who do not present and whose needs may be as great; by following the course of remission and relapse, adjustment and disability in defined populations. Follow-up of cohorts is necessary to detect early sub-clinical and perhaps reversible disease and to discover precursor abnormalities during the pathogenesis, which may offer opportunities for prevention.

7) To search for *causes of health and disease* by computing the experience of groups defined by their composition, inheritance and experience, their behaviour and environments. To confirm particular causes of the chronic diseases and the patterns of multiple causes, describing their mode of operation singly and together, and to assess their importance in terms of the relative risks of those exposed. Postulated causes will often be tested in naturally occurring experiments of opportunity and sometimes by planned experiments.

Having reviewed the uses of epidemiology, next we shall focus on the developments in modern epidemiology.

## 1.2.3 Developments in Modern Epidemiology

In a sense epidemiology is as old as medicine itself. *Hippocrates* considered the father of modern medicine first suggested that the development of human disease might be related to the external, as well as, personal environment of an individual.

Since then, for approximately 2000 years, not much attempt was made to measure impact of "causes" of disease.

For the 1$^{st}$ time in 1662, the patterns of disease in a population (in London) were quantified by *John Graint*. His recognition of the value of routinely collected date in providing information about human illness forms the basis of modern epidemiology.

The developments in epidemiology are traced briefly for your reference:

In 1850's *John Snow* formulated and tested a hypothesis concerning origins of Cholera epidemic in London.

Since World War II there has been rapid and systematic progress in development of both principles and methods of epidemiologic research.

When studying about epidemiology you will find that the first major development was the design of studies and techniques for collecting and analyzing data in order to evaluate risk factors for chronic disease. One such strategy to address the problem of long latency periods was to get together a group of individual with a particular disease and a comparable group of persons without it. Information about their previous medical history, health habits was collected. This approach (*case-control study*) allowed the investigators to look back in time (retrospectively) to assess quickly the effect to exposure to a given factor without having to wait for 10-20 years for the disease to develop. The association between cigarette smoking and lung cancer was shown through a classic study by *Doll* and *Hill* in 1950. This approach became especially relevant because morbidity and mortality due to infectious disease, famine and pestilence were decreasing in the countries of the developed world e.g. the U.S. With industrialization there were improvements in nutrition, housing, sanitation, water supply, treatment methods including use of antibiotics and immunization.

Simultaneously there has been emergence of chronic diseases, which characteristically have latency periods of 10-20 years and are now major causes of mortality. The change in mortality patterns has had far-reaching implication. The case-control approach did not help scientists to address all problems. There was still a need for obtaining accurate and precise information about the period of exposure, a risk factor or causative agent prior to the occurrence of an event. This problem was solved by using *cohort study*. Here a group of individuals without the disease of interest were taken, classified with respect to their exposure status at the start of the study and then subsequent development of the disease was monitored in exposed, as well as, non-exposed subjects over time. A very good example of an early study is the *Framingham Heart Study*. Here 5200 residents of Framingham, Massachusetts in the U.S. have been followed up for more than 35 years to explore the relationship of a wide variety of risk factors such as lipoproteins, cholesterol etc. with coronary heart disease.

Later epidemiologic principles and methods have been applied to the design conduct and analysis of clinical trials (intervention studies, experimental studies). In these studies, the investigators themselves allocate to participants the exposures being studied. These studies recognize the potential importance and are based on manipulating the human environment just like basic researchers control conditions in the laboratory.

The first formal human experiment was the field trail of the polio vaccine in the 1950's. Now clinical trails are an integral part of the evaluation of new preventive therapeutic agents and procedures.

Computer technology is now used for storage and analysis of data since tens of thousands of individuals may be studied in addition to information being recorded for a large number of variables. Basic research adds to our biologic understanding of why an exposure causes or prevents disease. But epidemiology allows the quantification of the magnitude of the exposure-disease relationship in humans. Research has often provided information based on which public health decisions have been made to alter the risk through intervention.

Several research studies, which are well known and well publicized in the media, are:

- The effect of oral anti-diabetic medication
- Saccharin ionizing radiation and leukemia
- AIDS
- Coffee drinking and pancreatic cancer
- Blood pressure and stroke
- Ascorbic acid and common cold
- Aspirin and myocardial infarction
- The MRFIT research
- Dietary fat and risk of breast cancer
- β-carotene and cancer
- Vitamins, folic acid and neural tube defects

From our discussion presented above, you must have now got a brief review of the development in modern epidemiology. While reading about the developments you would have realized that epidemiology involves a study of the various variables in the disease-exposure relationship in the humans. The next sub-section highlights the various descriptive variables studied while undertaking research.

### 1.2.4 Descriptive Variables for the Health of the Community

In all researches, the scientist studies variables. What do we mean by variables? Generally speaking, a characteristic that tends to vary from subject to subject or from unit to unit is called a variable. These could include the demographic and social variables, health-related outcome variables, variables related to community infrastructure. Let us get to know these variables.

*Demographic and social variables*

The demographic and social variables includes:

- Age and sex distribution
- Socio-economic status
- Family structure, including marital status and number of single-parent families
- Racial, ethnic and religious composition etc.

*Variables related to community infrastructures*

Variables related to community infrastructure would include:

- Availability of social and health services
- Quality of housing stock
- Social stability (residential mobility) etc.

*Health-related outcome variables*

The variables depicting health-related outcome may include:

- Homicide and suicide rates
- Infant mortality rate
- Mortality from selected conditions (cause specific)
- Magnitude of chronic and infectious diseases
- Alcoholism and drug abuse rates
- Teenage pregnancy rates
- Birth rate etc.

Generally, variables are broadly classified as follows:

Qunatitative and Qualifative variables

Discreet and Continuous variables, and

Stochastic and Deterministic variables

A brief review of these categories follows.

*Qualitative and Quantitative Variables* : We know that characteristics such as birth weight, body temperature, height etc. have a numeric outcome for example the normal birth weight is considered 2.5 kg, the normal body temperature 98.7°F, or respiration rate 68 per minute etc. When such numeric assignment is done, the characteristic becomes a quantitative variable in true sense. Use of quantitative variables such as body temperature, birth weight, height etc. is of course common but, at some stage in patient management, they tend to be interpreted as "qualities" such as high/normal/low and preterm/term. The clinical interpretability becomes easier by assigning such "qualities". For example, malnutrition of a child can be classified as none, mild, modrate and severe category, body temperature as high/normal/low etc. Thus we often use the term 'quantitative' and 'qualitative' to distinguish between these variables.

*Discrete and Continuous Variables*: The variable that can take only finite, generally small, number of values in a range is called *discrete*. Only those variables which can take a small number of values, say, less than 10, are generally considered discrete. Variables such as Apgar score, blood group and birth order are some examples of a discrete variable. A variable which can take infinite number of values within a range is called *continuous*. Anthropometric measurements such as weight, height and mid-arm circumference, laboratory measurements such as iodine level, Hb level and serum bilirubin level are examples of a continuous variable.

*Stochastic and Deterministic Variables*: There are measurements that are considered known for the subjects. The others are those that are susbequently obtained. The former are factors and the latter responses. Let us understand this with the help of an example. While studying the relationship between maternal anaemia and birthweight, anaemia is a factor and the outcome (birth weight) is a response. Responses are considered *stochastic* because they are subject to chance fluctuations. They can not be exactly predicted. To illustrate in our example above birthweight (response) is likely to be influenced/effected by anaemia. Factors are considered *deterministic* because they are known before hand and are no longer subject to chance fluctuations. Most of the inferential statistical methods we discuss in these articles apply to the stochastic rather than to the deterministic variables.

Many a times we are not certain of "causal" factors, then we refer to an exposure that is associated with a disease as a "risk factor".

There are three criteria for risk factors. These include:

i) The frequency of the disease varies by category or value of the factor e.g. cigarette smoking and lung cancer – heavy smokers are more likely to develop lung cancer than light smokers

ii) The factor must precede the onset of disease – this is particularly so for chronic diseases of long duration

iii) The observed association must not be due to any source of error.

With a review of the various descriptive variables which depict the health of the community, we end our study of epidemiology. We hope the discussion presented above would have provided you a good insight into what is epidemiology, its uses and purpose and the developments in modern epidemiology. Next, we move on to another concept i.e. the study of biostatistics.

## 1.3 BIOSTATISTICS

Statistics is an area of science concerned with the extraction of information from numerical data and its use in making inferences about a population from which the data are obtained. Biostatistics is the science of management of uncertainty in nutrition/ health research. The most common source of uncertainty in nutrition/health research is the natural biologic variability (variables such as age, sex, birth order, height, weight etc.) between and within individuals. Environmental variability (pollution, nutrition, exposure to microorganisms, insects, flies etc.) and variations between laboratories, instruments, observers, *etc.*, further aggravate these uncertainties. Other sources of uncertainties are incomplete information of the subjects, lack of medical knowledge, validity and reliability of the tools used for research etc.

*Statistics* deals with measuring these uncertainties, to evaluate their impact and, of course, to keep this impact under control. The general principles to minimize the impact of uncertainties are (i) inclusion of a control group where appropriate, (ii) proper selection of subjects, (iii) matching or randomization, and (iv) use of standardized methods. We will learn about these aspects later in this course. The statistician quantifies information, develops and uses various designs and sampling procedures, and is involved with analysis and inference making. She/he is also concerned with providing a quantitative measure of the goodness of the inference-making procedure.

Thus major contributions of statistics and statisticians are in designing experiments and surveys, thereby reducing there cost and size. When we predict we would like to know something about the error in our prediction. If we make decisions, we should know the chance that our decision is correct. We are unable to instinctively give answers to these questions. On the other hand, if we use scientific approach/ observations and apply statistical procedures in research we are able to provide the answers needed.

Statistics performs major roles in research. It helps us to assess random variation, to control confounding, to evaluate interactions between factors, to determine sample size, to decide on the designs for studies and facilitate decision making. This aspect will become clear as we read through this course on research methods and biostatistics. Most researches you will find include the scientific assessment of random variability, which we do by means of "significance testing or statistical hypothesis testing". What is a hypothesis? *Hypothesis is a tentative proposition suggested as a solution to a problem or as an explanation of some phenomenon.* We will get to know more about this in Unit 2 later in this course. You may have also read about "p values" and "statistically significant" while reading a research article. This has roots in the early years when statisticians pioneered the development of statistical theory. Their research problems were industrial and agricultural – typically they involved experiments that formed the basis for a choice between two or more alternative courses of action enabling a decision to be made.

Statistical hypothesis testing focuses on the *null hypothesis*, which is a hypothesis of no association between two variables. If the data provide evidence against the null hypothesis, then the hypothesis can be rejected in favour of *alternative hypothesis*, which very simply put states that there is an association. The alternative hypothesis can be one-sided. We can state that there is specifically a positive or negative association. We will learn more about this concept later in Unit 13.

With a brief understanding of biostatistics and its importance in research, let us next focus on understanding what is meant by scientific approach and what do we mean by the term research.

## 1.4 WHAT IS RESEARCH AND SCIENTIFIC APPROACH?

We have been using the word 'Research' frequently so far. What is *research*? In simple terms, we can say that *research is an endeavour to discover intellectual and practical answers to problems through the application of scientific methods to the knowable universe.* Thus in research, scientific methods and procedures have been developed and are used to answer questions like "what", "how" and "why" about phenomena of interest.

One may also explain the term as *"critical and exhaustive investigation or experimentation having as its aim the revision of accepted conclusions in the light of newly discovered facts"*. Thus researchers are constantly concerned not only with discovering new things but also verifying accepted conclusions or theories.

However, as *Pearson* has said "there is no shortcut to the truth .... No way to gain knowledge of the Universe except through the gateway of scientific method".

*Scientific method* and *scientific approach* in general are similar in all branches of science. Why? Because this approach and method is the approach and method of trained minds, we must remember that facts alone do not make a science, rather the method by which they are dealt with make a science. So what is science? Let us find out in the next sub-section.

## 1.4.1 What is Science?

"Science is an objective, logical and systematic method of analysis of phenomena devised to permit the accumulation of reliable knowledge. It is a systematized form of analysis" — "an intellectual construction" or "a working thought-model of the world and its aim is to describe and conceptualize the impersonal facts of experience in verifiable terms, as exactly and simply as possible, as well as, completely and meaningfully as possible".

*What do we mean by scientific method/research?*

Scientific method refers to a procedure or mode of investigation by which scientists acquire systematic, scientific knowledge. We can say that there are specific components in scientific methods which include:

1) Reliance on empirical evidence
2) Use of relevant concepts
3) Commitment to objectivity
4) Ethical neutrality
5) Generality
6) Predictions based on probability
7) Public methodology that affords testing of conclusions through replication

What do these components mean? Let us review them briefly.

1) *Reliance on empirical evidence*: Science and scientists are committed to the belief that "truth" or the source of knowledge is experience, wherein the scientists use reasoning so as to comprehend. Thus rational ideas are the guiding principle in making predictions that are then tested by scientists by observations.

2) *Use of relevant concepts*: Concepts are logical constructs or abstractions created from sense impressions, precepts and experiences. For example, stress is a concept. A psychologist may measure stress with a particular tool, in medicine we do the stress test for cardiovascular function, the biochemist may measure the extent of lipid peroxidation as a measure of oxidative stress, the food scientist may measure the amount of force (shear stress) required to break a gel. Scientists constantly rely on and use relevant concepts.

3) *Commitment to objectivity*: Science requires the scientist to be objective i.e. to set aside one's subjective (personal) considerations i.e. one's opinions, hopes or intuition. This means that the scientist must at all time "strive at self elimination in his judgments".

   Objectivity is a composite of :

   a) Repeated observations of a given phenomenon by the same observer will yield constant data (give the same results)

   b) Repeated observations by different observers will give constant data

4) *Ethical neutrality*: The scientist should be unbiased or unprejudiced. He should not be biased or influenced by appearances, have no favourite hypothesis or committed to an ideology.

5) *Generality*: In science our aims to discover the "thread of uniformity" which lies under the diversity we see. Once the scientist discovers uniformity, a generalization is formulated.

6) *Predictions based on probability*: Scientists believe that predictions about phenomena can be made only if we have a solid basis of trends that have been repeatedly observed and the probability that such trends can manifest in terms of concrete result. It would be well for us to remember what has been said "Scientific knowledge is a body of statements of varying degrees of certainty, some most unsure, some nearly sure none absolutely certain".

7) *Public methodology affording testing of conclusions through replication*: "Science is a public institution with a public methodology". Therefore every scientist is required to make known to others how (the methodology used) he/she arrived at the conclusions. Therefore, the methods and materials used should be made known and along with conclusions subject to critical scrutiny. Why? Because science method of knowing which is self-corrective in operation that learns from failures as from successes.

We must always remember that science is a collective and cooperative endeavour geared to the discovery of facts. Remember replicability is important; hence fellow scientists or critics would not be able to replicate a study and verify the conclusion unless the methodology is made publicly known. Also we should not think that replication is copying. *Replications* lend credence and add strength to the conclusions, if the conclusions arrived at are the same as before. If the conclusions are invalidated, remember that invalidation is an important contribution to science as is the verification of propositions.

Scientific research, therefore, is a "systematic and critical investigation about the natural phenomena to describe, explain and finally to understand the relations among them". The conceptual foundations of scientific research are elaborated next.

### 1.4.2 Conceptual Foundation of Scientific Research: Facts and Theory

Scientific research starts with facts and then moves towards theorising. To be useful, facts must be organized, and the primary purpose of the scientific method is to develop a mechanism of organizing the facts, as they accumulate, and become meaningful from the standpoint of their objectives. Through empirical investigations, scientists gather many facts. As these facts accumulate, there is a need for integration, organization, and classification in order to make the isolated findings meaningful.

When isolated facts are put in a perspective by integrating them into a conceptual scheme, which promote greater understanding, we approach the domain of science. When isolated facts are integrated into a conceptual scheme which promotes a better understanding of their nature and significance, it is clear that the facts have been already put in proper scientific perspective. Significant relationship in the data must be identified and explained. In other words, theories must be formulated. Theory may be defined as "a set of interrelated constructs (concepts), definitions and propositions that present a systematic view of a phenomena by specifying relations among variables, with the purpose of predicting and explaining the phenomena (Kerlinger, 1973).

Theory knits together the results of observations, enabling scientists to make general statements about variables and relationships among variables. For example, in Boyle's Law, a familiar generalization summarizes the observed effects of change(s) in temperature on the volumes of all gases by the statements – "When pressure is held constant, as a temperature of a gas increases, its volume is increased and as a temperature of a gas is decreased its volume is decreased". This statement of theory not only summarizes previous information, but also predicts other phenomena by telling us what to expect of any gas under any change(s) in temperature.

In spite of the strong case that has been made for the role of theory in research, it will be appreciated that a theory has to be amended or abandoned when the discovery of new facts can no longer accommodate it. Alternatively, it may be subsumed by a wider, more embracing theory when it is realized that the situation which is contained by the theory is one instance of a more general case. Theories generated by the means that we have indicated, do not lead to 'eternal truth'; rather, they should be looked upon as useful conceptual frameworks which are adequate for present purpose or a given situation. Thus every theory is subject to modification as and when we get new facts and evidence that contradict the generalizations made earlier on.

### *Purpose of Theory*

There are several purposes to be served by a theory in the development of science. We shall briefly consider three of them here. First, theory summarizes and puts in order the existing knowledge in a particular area. It permits deeper understanding of data and translates empirical findings into a more easily retainable and adaptable form. The theory of oxidation for instance, places into focus many of the chemical reactions common to everyday life.

Secondly, theory provides a provisional explanation for observed events and relationships. It identifies the variables that are related and the nature of their relationships. A theory of learning, for example, could explain the relationship between the speed and efficiency of learning and such other variables as motivations, reward and practice.

Lastly, theory permits the prediction of the occurrence of phenomena and enables the investigator to postulate and, eventually, to discover hitherto unkown phenomena. At the time when the 'Periodic Table' was being completed, for instance, certain gaps were noted in the sequence of the elements. Since theory provides that, there should have been no gaps, scientists were spurred on to look for the other missing elements. In time, these were found, anticipated by theory. Theory, therefore, stimulates the development of new knowledge by providing the lead for further inquiry.

Finally a word about how to develop a theory.

### *Developing a Theory*

It is important to stress that good theories are not born out of imagination, they do not originate merely through arm chair reflection. A theory is built upon collected facts. The investigator then searches, makes intelligent guesses as to how the facts are ordered, adds missing ideas or links, and puts forwards a hypothesis; deduces what consequence should follow from the hypothesis and looks for further facts which are consistent or otherwise with the deductions; builds a wider generalization or conceptual framework on more facts; and eventually outlines a theory. Theories are solidly based on evidence. And they are important practical tools which enable us to advance our knowledge still further. Once a theoretical framework has been elaborated we know what facts to look for to confirm or to deny the theory; also, we have a conceptual framework inside with which our evidence can be tested.

Theories always involve terms that refer to matters that cannot be directly observed. For example, gravity itself cannot be directly observed, though the effects of gravity can be. Gravity and gravitation are both theoretical terms. The terms of a theory or theoretical statement are sometimes referred to as constructs. Thus, many theories of learning refer to a motivational factor in behaviour. Now motivation is not directly observable. It is a theoretical term; or, we may refer to it as a construct. The term implies that it is a construction of the scientist's imagination.

Now that we have a clear insight into what is meant by scientific approach and its conceptual foundation, let us next, dwell on what is research.

### 1.4.3 What is Research?

One may define research as *"a critical and exhaustive investigation or experimentation having as its aim the revision of accepted conclusions in the light of newly discovered facts"*.

It has also be defined as *"the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art"*.

Let us discuss each part of this definition, which may help to clarify the definition for you.

a) *Manipulation of things, concepts or symbols*: In research we deal with things very often e.g. in nutrition many experiments are done on rats – they may be given a very specific diet, sometimes they may be starved. In essence we handle them in a specific way (purposeful handling) i.e. manipulation, which is an aspect of experimentation. Science also deals with concepts (the terms which designate the things/abstract/notions about which a science tries to make sense are concepts) e.g. psychologists deal with anger, emotion, problem-solving potential, fear, satisfaction etc.

b) *For the purpose of generalizing*: In research we purposeful manipulate/control things or concepts in order to arrive at statements of generality i.e. after conducting the research/study we should be able to arrive at conclusions which are applicable to the population i.e. extend beyond the sample to the population segment we have chosen to study. For example, we find that a drug helps to reduce blood glucose levels in an experiment with a group of Type II diabetes. Remember however that the conclusion tells us that we can expect "something" in a class of things under a matter of degree/level – it can be high or low – and various conclusions therefore will vary in their degree of generality.

c) *To extend, correct or verify knowledge*: When a generalization is drawn it contributes to the body of knowledge, which was existing at that point of time. For example, when *Jenkins* did his work on Glycemic Index in the 70's – this was added to the knowledge about how foods may be distributed in the diet of diabetics and how one can help to control blood glucose levels.

   Some generalization may make scientists re-examine existing knowledge or make us amend or modify our views. For example, in Physics newer discoveries about subatomic particles led to the revision of accepted knowledge about matter.

   Generalizations can be used to better understand phenomena in depth. Some generalization may show that the phenomenon under inquiry is governed by the same law applicable to certain other phenomena. This will help us to appreciate that there are linkages between them.

   Thus we can extend knowledge by research, on the other hand, the findings may help to correct inconsistencies or errors in the existing body of knowledge. An important aspect of scientific activity is verification of conclusions, which have found a place in the established system of knowledge.

d) *Knowledge may be used for construction of a theory or practice of an art.* What does this mean? It implies that extended, corrected and verified knowledge has two possible uses (which are not mutually exclusive: (i) theoretical, and (ii) practical.

   Many a time's research is undertaken for the sake of knowledge. Such use is "theory-oriented" as is often known as *'pure', basic'*, or *'theoretical'* research. On the other hand it may not be an end in itself but we may use it for the welfare of people – these are *"applied"*, *"action oriented"*, *"practice oriented"*.

In nutrition we undertake study of various kinds. They can be *pure,* as well as, *applied* and *action oriented.* We will learn more and in details about the types of research study designs used in nutritional epidemiological research in the subsequent units. Here, a brief review of the types of research is presented with the main focus on the scope of research in nutrition.

## 1.4.4 Scope of Research in Nutrition

In nutrition research, our focus is basically on the relationship between diet, nutrients and health. Many of the studies of nutrients and their functions especially that done in the 1900's extended and continues to extend the body of knowledge.

We do undertake experiments in both animals and humans. The classic methods of nutrition studies included basic biochemistry, animal experimentation and metabolic studies. But these studies do not fully help to explain the relation between diet and disease. Hence, epidemiologic studies are undertaken. But studying diets is a complex exercise. One can easily quantify the number of cigarette smoking. In contrast, diet represents a complex array of factors that are strongly interrelated. With few exceptions all individuals are exposed to fat, fibre, protein (food consumed) etc. Therefore, we cannot say the factor is either absent or present in the diet. Generally there is a range (continuum) of intake, with variations being limited to some extent. Also most individuals more often than not do not make clear changes in their diet at identifiable points in time. Typically a person's dietary pattern evolves over years. Moreover, persons are generally not aware or knowledgeable about the portions, size or the content of the food.

There is substantial (biologically meaningful) variation between persons for most nutrients. In fact without such variation, observational studies of individuals would not be possible.

In nutritional epidemiology, the concepts, hypothesis and techniques are derived from a variety of sources including metabolic and biochemical studies among humans, animal experiments and in vitro studies. Various types of researches are undertaken such as:

- *Correlation or Ecological Studies*: Wherein disease rates in population have been compared e.g. fat intake and cancer incidence in populations from different countries which has been derived from food disappearance data. However, a major problem is that there can be numerous other potential factors, which may affect the disease other than the dietary factor.

  International correlation studies cannot be independently reproduced which is an important part of the scientific process.

- *Special Exposure Groups*: Groups within a population that consume different diet are often studied e.g. in case of colon cancer studies were conducted on Seventh Day Adventists, a section that is largely vegetarian. However, these studies suffer from limitation similar to the correlation studies. There can be many possible alternative factors, which lead to specific disease patterns in such special groups.

- *Migrant Studies and Secular Trends*: You must have encountered studies on Japanese immigrants to US and Asian immigrants. Migrant studies are useful when the correlations that we observe in ecologic studies are because of heredity. In case of cancer, it has been observed that the migrant population acquires rates of incidence, which are characteristic of the new geographic region they have settled in.

  These groups can also be studied to examine the relevant time of exposure (latency). Scientists have observed that with time there have been major changes in the rates of a disease and provided evidence that more than heredity other factors like diet and lifestyle have very important roles.

- *Case-Control and Cohort Studies*: In *case-control studies* information about previous diet of patients is gathered and this compared with individuals (subjects) free from the disease (the control group). In *cohort studies* disease free individuals are followed up over time to determine the rates vis-à-vis the dietary component(s) of interest to the investigator.

  As you would realize, case-control studies take less time since we require smaller number of subjects and we do not need to follow them up. However, we are not very sure whether we can consistently get the same results. This is because of the methodology (bias) and because ultimately there is a limited range in the diet, as well as, the inherent errors in estimating dietary intakes (since intakes are based on recall). Cohort studies have an advantage because dietary information is collected before diagnosis of disease, illness will not affect recall. The difficulty is that the scientists must study very large number of subjects.

- *Controlled Trails*: In terms of testing a hypothesis, this design is much more rigorous because variables which might influence the phenomenon are taken care of (controlled) by randomization. Such experiments are very useful for examining the role of minor dietary components, which can be, administered in controlled, measured amounts in the form of pills and capsules e.g. minerals, trace elements and vitamins.

  However there are limitations because scientists may involve relatively smaller number of subjects, there may be non-compliance, there can be spontaneous changes in the diet.

So what do we look at in these studies? We may like to look at:

a) The strength of association

b) The consistency of results in various studies and populations

c) Is there a dose-response relationship

d) Is it a time-related (temporal) relationship

e) The biologic plausibility

But we find more often than not that relationships between diet and disease are very complex. The type of foods eaten and their amounts can be influenced by non-dietary factors, which also influence disease rates e.g. age, exercise, and smoking. These factors can modify and confound relationships.

Further there are inter-relationships between nutrients and thus one nutrient can interfere (like iron and zinc) or enhance (like vitamin C and iron) the absorption of another or influence the metabolism.

Because of this biologic complexity (both in terms of human metabolism and food composition) and variations in food eaten on a day-to-day basis, cultural and religious practices e.g. fasting, seasonal effects, food choices and error in measurement of food intake, it is not possible to predict with certainty the health effects of any food solely on the basis of its content of one specific factor.

The degree of random variation differs according to a nutrient. For macronutrients, which contribute to energy (Kcal) intake, the degree of variation is comparatively less than micronutrients which are not present in all foods.

Thus there are large variations between persons and within persons. It is not possible to measure dietary intake for large number of subjects. But if only one or a few days are measured, a subject's true long-term intake is not likely to be truly represented.

The areas in which nutrition/health research is undertaken is vast in various age-sex-lifestyle groups (both healthy and with disease). Areas in which nutrition/health research is increasingly undertaken are highlighted next.

- Functions of nutrients at cellular level (effects of deficiency and excess)
  - Nutrients and cell function
  - Nutrients and cell signaling
  - Nutrients and genetics
  - Nutrigenomics
  - Nutrient inter-relationships
- Factors influencing absorption of nutrients
- Nutrition and development
- Food components (including phytochemicals) and disease from preventive, promotive and therapeutic perspectives
- Rapid appraisal methods
- Validation studies
- Biochemical markers

It will be worthwhile if you can go through at least five highly read journals both at the international and national level to get an idea about the kind of research being undertaken in the area of nutrition and health. To name a few among the international ones, you could review the American Journal of Clinical Nutrition, Journal of the American Dietetic Association, European Journal of Clinical Nutrition, the Lancet, and Journal of Nutrition etc. Among the National Journals, Indian Pediatrics, Indian Journal of Medical Research, Indian Journal of Community Medicine etc.

We will learn more and in details about the types of research study designs used in nutritional epidemiological research in the subsequent units. We end our study of the basic concepts in epidemiology and research here, with a brief review of the research process in general.

### 1.4.5 The Research Process

The research process is the paradigm of research project. In a research project, researcher engages himself/herself in various scientific activities in order to produce knowledge. Although each research project is unique in some ways, all projects, regardless of the phenomenon being studied, involve by and large, some common activities. Each of these activities are interdependent. The research process is a system of these interrelated activities. The various activities are conveniently grouped into six stages as shown below:

| | | |
|---|---|---|
| Stage I | : | Selection and Formulation of a Problem |
| Stage II | : | Formulation of Hypothesis |
| Stage III | : | Formulation of Research Design |
| Stage IV | : | Collection of Data |
| Stage V | : | Analysis and Interpretation of Data |
| Stage VI | : | Generalization |

The stages of research are interdependent. The researcher usually enters the research process at stage I. However, when he/she enters second stage, he/she has to draw on past studies to formulate his/her hypothesis. Similarly, to select a research design the researcher has to keep in mind the problem and the hypothesis. A researcher, who has no knowledge of how to collect and analyze data, may find himself/herself unable to formulate a testable hypothesis, or formulate the research design. This brief discussion on the research process makes it very clear that each of these six stages of research process is dependent upon others.

The research process is also cyclic in nature (as shown in Figure 1.2). In fact, the research process is not complete at the stage VI. The process leads to two situations: The first situation may be that the data did not support or partially support the hypothesis. In this situation, that is even if the research is successful and the findings of stage VI confirm the hypothesis of stage II, it is advisable to repeat the study preferably with a different sample as to reconfirm the findings. This will also support the contention that the hypothesis cannot be rejected. The exact repetition of a study is called *replication*. This is an important cardinal principle of research; particularly experimental research. Replication is the process of repeating the same treatment/ experiment on more than one testing unit.

It provides data to quantify the experimental error and helps to reduce it. It increases the confidence in the results and help in brigning the clear signals to the fore.

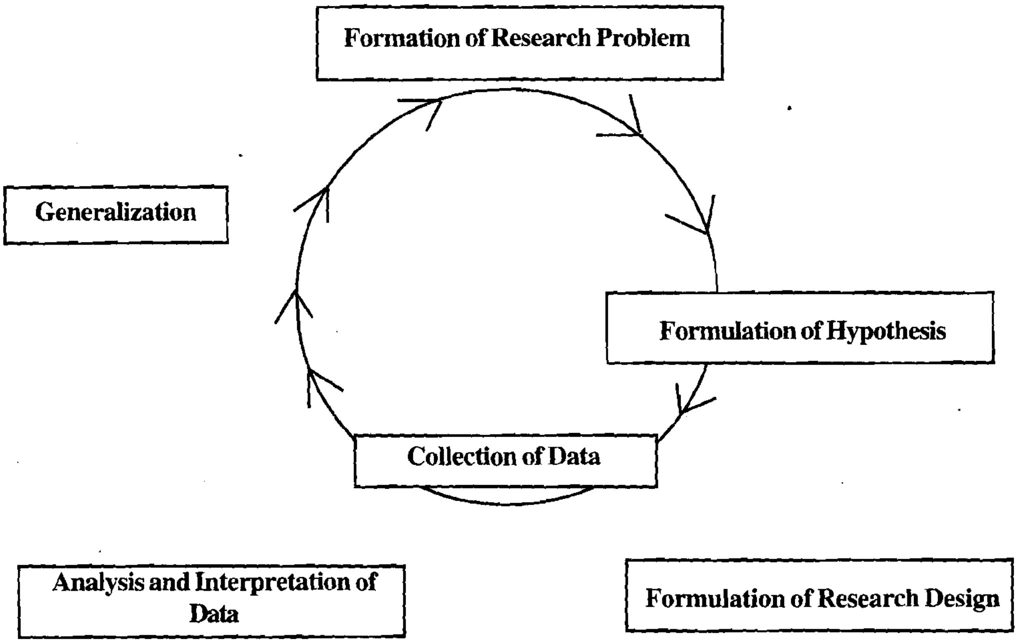


**Figure 1.2: Stages of the research process**

Another characteristic feature of the research process is 'self correction'. In a situation, when the data did not support or only partially support the hypothesis and the researcher has sufficient reasons to decide that the hypothesis is adequate then he/she may decide that the failure to confirm the hypothesis is due to error in selecting a sampling design or in measurement of the key concepts or in analysis of data. In these situations the researcher may decide to repeat the study beginning with the faulty stage after rectifying the faults. Finally, the six stages of the research process make the study potentially replicable. The researcher designs his/her study in such a way that either the researcher or others can replicate it. The replication of study substantiates the fact further that the findings are not due to mere coincidence.

To help recapitulate what you have learnt so far, next we have included few exercises in the check your progress exercise 1. Answer these questions.

**Check Your Progress Exercise 1**

1) Define the following

   a) Epidemiology ..........................................................................

   ..........................................................................

   b) Nutritional epidemiology ..........................................................................

   ..........................................................................

c) Research: ..........................................................................................

..........................................................................................................

2) List the important purpose(s) of epidemiology.

..........................................................................................................

..........................................................................................................

..........................................................................................................

3) What are the specific components in scientific methods?

..........................................................................................................

..........................................................................................................

..........................................................................................................

4) Visit any library in the area of nutrition/health, which has research journals. Specifically if you can pick up any volume of one of the following:

   a) American Journal of Clinical Nutrition
   b) European Journal of Clinical Nutrition
   c) Journal of the American Dietetic Association
   d) Journal of Nutrition

Now pick an article that deals with experimentation, either in the laboratory or a study involving a sample survey. The article may not contain the actual data collected in the experiment.

1) Discuss the research objective
2) Identify the experimental units, the sample and the population

## 1.5 LET US SUM UP

In this unit the basic concepts, constructs related to research and epidemiology were presented. We learnt that epidemiology is the study of the distribution and determinants of disease frequency in human population. The aim of epidemiology is to explain the etiology, predict and prevent occurrence of disease. The various uses of epidemiology were described and the descriptive variables studied in epidemiology were highlighted.

The unit also described the concept of science and the scientific approach and methods used in research, particularly with reference to nutritional epidemiology, and highlighted that the concepts, hypothesis and techniques in nutritional epidemiology are derived from a variety of sources including metabolic and biochemical studies among humans, animal experiments and in vitro studies. Various types of researches such as case-control, cohort, correlational studies etc., undertaken in nutritional epidemiology were further elaborated. Areas in which nutrition/health research is increasingly undertaken were also highlighted.

## 1.6 GLOSSARY

**Glycemic Index** : the glycemic index measures how fast and how much a food raises blood glucose levels. Foods with higher index values raise blood sugar more rapidly than foods with lower glycemic index values do.

**Nutrigenomics** : the study of how different foods may interact with specific genes to increase the risk of common chronic diseases such as type 2 diabetes, obesity, heart disease, stroke and certain cancers.

**Phytochemicals** : they are usually used to refer to compounds found in plants that are not required for normal functioning of the body but that nonetheless have a beneficial effect on health or an active role in the amelioration of disease. Examples of phytochemicals include lutein found in corn yellow, lycopene in red tomatoes, anthocynin in blueberries etc.

## 1.7 ANSWERS TO CHECK YOUR PROGRESS EXERCISES

**Check Your Progress Exercise 1**

1) a) Epidemiology is the simplest and most direct method of studying the cause of disease in humans. It is a discipline that describes, quantifies and postulates causal mechanisms for health phenomena.

   b) Nutritional epidemiology deals with the study of the relationship, if any, between nutrient/dietary intake and disease condition.

   c) Research may be defined as the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art.

2) The purpose of epidemiology is:
   - To study the history of the health of populations
   - To diagnose the health of the community
   - To study the working of health services with a view to their improvement.
   - To *estimate* from the group experience what are *the individual risks* on average of disease, accident and defect, and the chances of avoiding them.
   - To identify syndromes by describing the distribution and association of clinical phenomena in the population.
   - To complete the clinical picture of chronic diseases and describe their natural history.
   - To search for causes of health and disease by computing the experience of groups defined by their composition, inheritance and experience, their behaviour and environments.

3) The specific components in scientific methods are:
   - Reliance on empirical evidence
   - Use of relevant concepts
   - Commitment to objectivity
   - Ethical neutrality
   - Generality
   - Predictions based on probability
   - Public methodology that affords testing of conclusions through replication

4) Carry out the assignment and answer based on your own understanding.