HEALTH RESEARCH METHODOLOGY
A Guide for Training in Research Methods

Second Edition

WORLD HEALTH ORGANIZATION
Regional Office for the Western Pacific
Manila, 2001

CONTENTS

FOREWORD

Scientific research plays a very important role in our efforts to maintain health and combating diseases. Research helps us create new knowledge and develop proper tools for the use of existing knowledge. Not only does it enable health care providers to diagnose and treat diseases, research also provides evidence for policies and decisions on health and development.

WHO and its Member States are aware of the importance of research. However, health research has not been a priority in many developing countries in the Region. The lack of research methodology and the absence of qualified researchers hinder many developing countries to conduct health research by themselves. In many countries, the system for management and coordination of health research has not been established or is not functioning properly.

WHO is committed to stimulating scientific research in developing countries. An articulate and clearly defined WHO framework and vision on research and partnership with Member States will strengthen research capacity in developing countries. The WHO Regional Office for the Western Pacific has organized more than 20 training courses on health research design and methodology in the last two decades. In 1992, the Regional Office published a training manual entitled *Health Research Methodology: A Guide for Training in Research Methods*. Since then, the manual, well received by readers worldwide, has been translated into Chinese, Khmer, Laotian, Mongolian and Vietnamese.

To accommodate requests from readers to incorporate recent developments on research methodology and experiences of past training courses, the manual has been revised and reissued.

We hope this revised version of the landmark manual will help scientists, researchers, health practitioners and administrators to learn and practise the concepts and principles of scientific research. The knowledge of the scientific methods will help them design and conduct research projects with precision in their own countries. The publication of the revised manual also reiterates our commitment to developing countries in the Region to help them build and strengthen the health research systems.

Shigeru Omi, MD, Ph.D.
Regional Director
WHO Western Pacific Regional Office

INTRODUCTION

This is a revised version of an earlier manual on Health Research Methodology and deals with the basic concepts and principles of scientific research methods with particular attention to research in the health field.

The research process is the cornerstone for informed and effective decision-making, and is integral to countries' efforts to improve the health of their populations and the effectiveness of their health systems, particularly during times of dramatic epidemiological, demographic, and economic changes that profoundly affect health systems. Research on (1) health policy and health systems, (2) epidemiology dealing with noncommunicable diseases and existing, new and emerging communicable diseases, (3) reproduction, child health and nutrition, including domestic or sexual violence, and (4) social-behaviour, including analysis of peoples' health seeking processes and their beliefs, knowledge and practices about health and illness, conducted by multi-disciplinary teams will enhance developing countries' efforts to fight diseases and maintain health for the public.

The manual describes methods for planning and conducting scientific research: from formulation of problems to setting research objectives, to designing the study, including methods of data collection, statistical analysis as well as interpretation and dissemination of the results. The earlier manual, used as resource and guide for the conduct of workshops on health research methodology in various countries of the Western Pacific Region has been expanded to include more details on some of the commonly used statistical methods and to clarify the points raised during workshops. The discussion on biases has been expanded considerably.

This manual is expected to be used by the WHO Western Pacific Regional Office as a reference guide in training young scientists to conduct health sciences research. It will be used as a starting point and not as a comprehensive textbook on research methods. Many excellent textbooks are available for this purpose and are referenced in the manual. We have tried to use real life examples from the Region for illustrating the principles and methods used in the manual to make it more relevant to the regional context.

The manual will be useful in planning a research project, especially in preparing a research grant application for a donor agency. In particular, the attached copy of the application form of WHO serves as a guide. The issues discussed in the manual will help the researcher to focus on issues of importance before the study is proposed and undertaken. In addition, the manual would also be useful when writing a thesis to meet academic requirements of a degree in the health field.

We hope that this manual will not only provide basic information on research methods in the health field, but also stimulate the reader to inquire further into the complex area of research methodology as well as increase the productivity of the young researcher in the Region. We hope it will attract researchers to conduct further studies in the health field, be it a clinical trial or field epidemiology or study of health services.

ACKNOWLEDGEMENTS

Chapter 1

Research and Scientific Methods

1.1 Definition

Research is a quest for knowledge through diligent search or investigation or experimentation aimed at the discovery and interpretation of new knowledge. Scientific method is a systematic body of procedures and techniques applied in carrying out investigation or experimentation targeted at obtaining new knowledge. In the context of this manual, research and scientific methods may be considered a course of critical inquiry leading to the discovery of fact or information which increases our understanding of human health and disease.

1.2 Categories of research

1. Empirical and theoretical research

The philosophical approach to research is basically of two types: empirical and theoretical. Health research mainly follows the empirical approach, i.e. it is based upon observation and experience more than upon theory and abstraction. Epidemiological research, for example, depends upon the systematic collection of observations on the health-related phenomena of interest in defined populations. Moreover, even in abstraction with mathematical models, advances in understanding of disease occurrence and causation cannot be made without a comparison of the theoretical constructs with that which we actually observe in populations. Empirical and theoretical research complement each other in developing an understanding of the phenomena, in predicting future events, and in the prevention of events harmful to the general welfare of the population of interest.

Empirical research in the health sciences can be qualitative or quantitative in nature. Generally, health science research deals with information of a quantitative nature, and this manual deals exclusively with this type of research. For the most part, this involves the identification of the population of interest, the characteristics (variables) of the individuals (units) in the population, and the study of the variability of these characteristics among the individuals in the population. Thus the quantification in empirical research is achieved by three related numerical procedures: (a) measurement of variables; (b) estimation of population parameters (parameters of the probability distribution that captures the variability of observations in the population); and (c) statistical testing of hypotheses, or estimating the extent to which 'chance' alone may account for the variation among the individuals or groups under observation.

Taking chance, or probability into account is absolutely critical to biological research, and is the substance of research design. Research design, above all else, must account for and maintain the role of chance in order to ensure validity. It is statistical methods which preserve the laws of probability in our inquiry, and allow proper analysis and interpretation of results. Statistics are the tool that permits health research to be empirical rather than abstract; they allow us to confirm our findings by further observation and experiment.

2. Basic and applied

Research can be functionally divided into basic (or pure) research and applied research. Basic research is usually considered to involve a search for knowledge without a defined goal of utility or specific purpose. Applied research is problem-oriented, and is directed towards the solution of an existing problem. There is continuing controversy over the relative benefits and merits to society of basic and applied research. Some claim that science, which depends greatly on society for its support, should address itself directly to the solution of the relevant problems of man, while others argue that scientific inquiry is most productive when freely undertaken, and that the greatest advances in science have resulted from pure research. It is generally recognized that there needs to be a healthy balance between the two types of research, with the more affluent and technologically advanced societies able to support a greater proportion of basic research than those with fewer resources to spare.

3. Health research triangle

Yet another way of classifying health research, be it empirical or theoretical, basic or applied, is to describe it under three operational interlinked categories of biomedical, health services and behavioural research, the so-called health research triangle. Biomedical research deals primarily with basic research involving processes at the cellular level; health research deals with issues in the environment surrounding man, which promote changes at the cellular level; and behavioural research deals with the interaction of man and the environment in a manner reflecting the beliefs, attitudes and practices of the individual in society.

1.3   Scientific foundations of research

Several fundamental principles are used in scientific inquiry:

1. Order

The scientific method differs from 'common sense' in arriving at conclusions by employing an organized observation of entities or events which are classified or ordered on the basis of common properties and behaviours. It is this commonality of properties and behaviours that allows predictions, which, carried to the ultimate, become laws.

2. Inference and chance

Reasoning, or inference is the force of advances in research. In terms of logic, it means that a statement or conclusion ought to be accepted because one or more other statements or premises (evidence) are true. Inferential suppositions, presumptions or theories may be so developed, through careful construction, as to pose testable hypothesis. The testing of hypothesis is the basic method of advancing knowledge in science.

Two distinct approaches or arguments have evolved in the development of inferences: deductive and inductive. In deduction, the conclusion necessarily follows from the premises, as in syllogism (all A is B, all B is C, therefore all A is C) or in algebraic equations. Deduction can be distinguished by the fact that it moves from the general to the specific, and does not allow for the element of chance or uncertainty. Deductive inferences, therefore, are suited to theoretical research.

Health research, being primarily empirical, depends almost entirely upon inductive reasoning. The conclusion does not necessarily follow from the premises or evidence (facts). We can say only that the conclusion is more likely to be valid if the premises are true, i.e. there is a possibility that the premises may be true but the conclusions false. Chance must, therefore, be fully accounted for. Further, inductive reasoning is distinguished by the fact that it moves from the specific to the general – it builds.

3. Evaluation of probability

The critical requirement in the design of research, the one that ensures validity, is the evaluation of probability from beginning to end. The most salient elements of design, which are meant to ensure the integrity of probability and the prevention of bias, are: representative sampling, randomization in the selection of study groups, maintenance of comparison groups as controls, blinding of experiments and subjects, and the use of probability (statistical) methods in the analysis and interpretation of outcome.

Probability is a measure of the uncertainty or variability of the characteristic among individuals in the population. If the entire population is observed, the calculation of the relative frequencies of the variables provides all the information about the variability. If only a sample of individuals in the population is observed, the inference from the sample to the population (specific to the general) will involve the identification of the probabilities of the events being observed, as well as the laws of probability that allow us to measure the amount of uncertainty in our inferences. These objectives can be achieved only by the proper design of research which incorporates the laws of probability.

4. Hypothesis

Hypotheses are carefully constructed statements about a phenomenon in the population. The hypotheses may have been generated by deductive reasoning, or based on inductive reasoning from prior observations. One of the most useful tools of health research is the generation of hypotheses which, when tested, will lead to the identification of the most likely causes of disease or changes in the condition being observed. Although we cannot draw definite conclusions, or claim proof using the inductive method, we can come ever closer to the truth by knocking down existing hypotheses and replacing them with ones of greater plausibility.

In health research, hypotheses are often constructed and tested to identify causes of disease and to explain the distribution of disease in populations. Mill's canons of inductive reasoning are frequently utilized in the forming of hypotheses which relate association and causation. Briefly stated, these methods include:

(a)   method of difference – when the frequency of a disease is markedly dissimilar under two circumstances, and a factor can be identified in one circumstance and not the other, this factor, or its absence, may be the cause of the disease (for example, the difference in frequency of lung cancer in smokers and non-smokers);

(b)   method of agreement – if a factor, or its absence is common to a number of different circumstances that are found to be associated with the presence of a disease, that factor, or its absence may be causally associated with the disease (e.g. the occurrence of hepatitis A is associated with patient contact, crowding and poor sanitation and hygiene, each conducive to the transmission of the hepatitis virus);

(c)   the method of concomitant variation, or the dose response effect – the increasing expression of endemic goitre with decreasing levels of iodine in the diet, the increasing frequency of leukaemia with increasing radiation exposure, the increase in prevalence of elephantiasis in areas of increasing filarial endemicity, are each examples of this concomitant variation;

(d)   the method of analogy – the distribution and frequency of a disease or effect may be similar enough to that of some other disease to suggest commonality in cause (e.g. hepatitis B virus infection and cancer of the liver).

1.4   Study design

The epidemiological approach is based upon statistical principles in the structuring of research design. In this approach, research can be divided into that which is basically observational in type, and that which is experimental.

Observational types of studies generally employ the method of sample surveys, where a sample of the population is observed for various characteristics. This may be by actual interviews of the

subjects, by obtaining measurements of physical characteristics, or by simply extracting information from existing sources, such as disease registries, hospital or employment records. Surveys of the cross-sectional type (where the information on cause and effect is simultaneously gathered, and the time sequence cannot be determined) are considered to be hypothesis-generating studies, whereas surveys where the observations on cause and effect differ by way of a period of time (such as case-control studies and cohort studies) are considered to be analytical in nature, and inference of associations can be made.

Testing of hypotheses is best done by experiment, where all the factors other than those under consideration can be controlled. However, in human diseases, this is not often possible, due to ethical and practical considerations. Therefore, it is often replaced by so-called 'natural' experiments, or by carefully designed observational studies (case-control studies, cohort studies) with enough information about the 'extraneous' factors to be able to adjust for these factors in drawing inferences. These analytical observational studies can be retrospective (case-control) or prospective (cohort and retrospective cohort studies). These methods compare groups of individuals for differences in exposure or differences in outcome. They differ from experiments in that there is no direct intervention by the investigator, and the investigator cannot control the extraneous factors for any of the individuals under observation.

In either approach, statistical reasoning using the laws of probability guides the inferential process. Some basic assumptions are made about the population, its characteristics and the probability distribution, and the likelihood of the observations supporting or contradicting the stated hypothesis, is evaluated. Based on these calculated probabilities, the hypothesis is accepted or rejected (or the state of uncertainty is left unresolved, especially when the samples observed are too small for reliability). Specific study designs are discussed later in the manual.

The process of moving from hypothesis generation to hypothesis testing is illustrated below.

An observation, or series of observations triggers a hypothesis; a cross-sectional survey is undertaken to generate proper hypotheses; an observational study establishes associations and supports (or rejects) the hypothesis; and an experiment is conducted to test the hypothesis.

```
┌─────────────────┐
│   Case Series   │
└────────┬────────┘
         │
         ▼
┌─────────────────┐
│ Cross-sectional │
│ Surveys         │
└────────┬────────┘
         │
         ▼
┌─────────────────┐         ┌──────────────────────┐
│ Analytic Studies│────────▶│ Case-control         │
└────────┬────────┘         │ Cohort               │
         │                  │ Restrospective Cohort │
         │                  └──────────────────────┘
         │
         │                  ┌──────────────────────┐
         └─────────────────▶│ Experiment           │
                            │ Randomized Trials     │
                            │ Quasi-Experiment      │
                            └──────────────────────┘
```
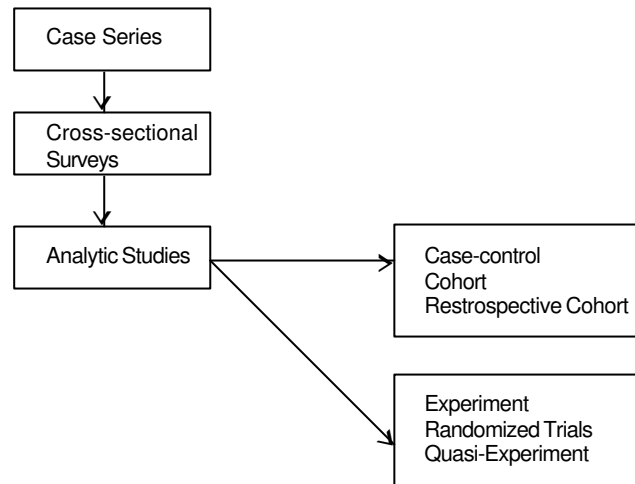
1.5   Planning and management of research

1. Research programme

As a complex activity, research requires careful planning, management and administration in its development and implementation. Within the constraints of the present world climate of restricted research budgets, it is becoming increasingly necessary that health research be programmed research, with clearly defined and practicably achievable objectives.

Some basic steps necessary in developing a research programme include:

(a)   defining the intended role and scope of the unit undertaking the research;

(b)   determining the capabilities and resources of the research unit, to include personnel, facilities, equipment, supplies, time and budget, and accessibility of research material;

(c)   selecting the research topic, considering factors such as

• magnitude of the problem and its impact;

• urgency of the need for a solution;

- relevance to the aims of the funding agency;

- amenability of the problem to investigation;

- feasibility of the approach;

- chances of success;

- expected impact of a successful outcome;

- spin-off in terms of training of staff and other research capability strengthening elements;

(d) constructing research protocols which will serve as the guiding documents for the execution, monitoring and evaluation of the research;

(e) setting up a well-defined administrative structure with lines of direction, supervision, consultation and collaboration based upon task-specific job descriptions;

(f) formulating a schedule of targets for consolidation of results and preparation of these results for dissemination, including publication in the scientific literature.

2. Execution of research

The mechanics of conducting research follow the simple steps of formulating the problem, planning the approach (research design) and executing activities within a strategic network leading to specific objectives which will give the solution to the problem. The following provides a framework for a research proposal into which the basic elements of a research study can be incorporated (these are discussed in more detail in Chapter 11):

a. Conceptualizing the problem:

- identifying the problem (what is the problem?);

- prioritizing the problem (why is this an important problem?);

- rationale (can the problem be solved, and what are the benefits to society if the problem is solved?);

b. Background:

   •   literature review (what do we already know?);

c. Formulating the objectives:

   •   framing the questions according to general and specific objectives;

   •   developing a testable hypothesis to achieve the objectives;

d. Research methodology:

   •   defining the population, characteristics of interest and probability distributions;

   •   type of study (observational or analytical, surveys or experiments);

   •   method of data collection, management and analysis:

      ◊   sample selection;

      ◊   measuring instruments (reliability and validity of instruments);

      ◊   training of interviewers;

      ◊   quality control of measurements;

      ◊   computerization, checking and validating measurements;

      ◊   the issue of missing observations;

      ◊   statistical summarization of information;

      ◊   testing of hypothesis;

      ◊   ethical considerations;

e. Workplan:

   •   personnel;

   •   timetable (who will do what, and when);

   •   project administration;

f. Plans for dissemination:

   •   presentation to authorities to implement the results of the research (if applicable);

- publication in scientific journals and other works (including those of the agency which funded the project) for wide distribution of the research findings.

A good proposal will also contain an executive summary giving an overview of the above topics in clear and simple language understandable by lay persons, and a list of references.

1.6   The research worker

Among the important qualities associated with successful research are:

- a spirit of adventure in seeking new facts;

- perseverance and patience;

- integrity to oneself and to the value of the scientific method;

    ◊   an analytical mind able to participate in critical thinking;

    ◊   receptivity to criticism at the professional level;

    ◊   openness of mind, and the ability to see the significance of the unexpected observation;

    ◊   objectivity.

1.7   Conclusion

Scientific inquiry is one of the most challenging enterprises of mankind, and the support that it receives is a measure of the strength, vitality and vision of a society. The approach and methods of research have slowly evolved to become ever more precise and efficient. The technology is at hand to explore the unknown. The success of this however, depends as ever on the individual and collective talents of the researchers bound by the tenets of science, such as those dealing with order, inference and chance, as accounted for and encompassed by solid research design and methodology.

Chapter 2

Research Strategies and Design

2.1   Introduction

The selection of a research strategy is the core of a research design and is probably the single most important decision the investigator has to make.   Therefore, the development of a research strategy is the main focus of this manual.   Essential components of the research design and the scientific basis for these will be discussed in the following chapters.

The research strategy must include the definition of the population of interest, the definition of variables (characteristics of the individuals in this population), their status and relationships to one another.   In testing a hypothesis, for example, an investigator may be able to assign the independence or exposure variable to a number of subjects in the study, and withhold it from others (controls) while controlling for other extraneous or confounding variables.   This strategy constitutes an experiment and covers hypothesis testing through intervention.

Another investigator may choose to compare people with and without exposure to a factor, and to analyse the incidence of a disease in these groups to find out if the disease is related to the exposure. This constitutes an analytical study, of which there are many varieties; this type of study also incorporates the testing of hypotheses.   Still another investigator may simply describe the distribution of a phenomenon or the outcome of a programme.   This constitutes a descriptive study with no intervention and no prior hypothesis.

In all of the above situations, observations are made on a group of people, and inferences are made about relationships or associations of various 'exposures' to 'outcomes'. The inferences reached are always subject to uncertainty due to the variation of characteristics across the population. The accuracy of the inference depends, therefore, on the accuracy of the information collected and on the representativeness of the subjects observed to the larger group of subjects in the population, as well as on the accuracy of the statistical methods used to draw the inference. In order to develop a good research strategy, we need to understand the nature of these 'errors' or 'variations' and the methods available to measure the errors.

2.2   Errors in the inference

Two common sources of error that need to be controlled result from problems with 'reliability' and 'validity'. Our inference should have high reliability (if the observations are repeated under similar conditions, the inferences should be similar) and high validity (the inference should be a reflection of the true nature of the relationship). The reliability and validity of inferences depend on the reliability and validity of the measurements (are we measuring the right thing, and with accuracy?) as well as the reliability and validity of the samples chosen (have we got a true representation of the population that we are drawing inferences from?). The reliability of a sample is achieved by selecting a large sample, and the validity is achieved by ensuring the sample selection is unbiased. In statistical terms, reliability is measured using 'random error' and validity by 'bias'.

2.2.1 Reliability

Reliability of measurements

If repeated measurements of a characteristic in the same individual under identical conditions produce similar results, we would say that the measurement is reliable. If independent, repeated observations are taken and the probability distribution is identified, the standard deviation of the observations provide a measure of reliability. If the measurement has high reliability, the standard deviation should be smaller. One way to increase the reliability is to take the average of a number of observations (the average having a smaller standard deviation – known as standard error of the mean [sem] – than the standard deviation of the individual observations).

Reliability of study

> A result is said to be reliable if the same result is obtained when the study is repeated under the same conditions. The natural variability in observations among individuals in the population is commonly known as random error. For example, if one is measuring the systolic blood pressure (SBP) of individuals, it has been observed that the measurements in large groups of people would follow a 'normal' distribution, so that the standard deviation of SBP is used as a measure of random error in SBP measurements. Clearly, if the standard deviation is small, repeated studies from this population are bound to come up with similar results. If the standard deviation is large, different samples from the same population will tend to differ substantially. Since we are often dealing with summary measures from samples that have standard deviations inversely proportional to the square root of the sample size, increasing the sample size increases the reliability of these measures. (More on these issues in Chapter 5.)

2.2.2 Validity

A measurement is said to be valid if it measures what it is supposed to. If a measurement is not valid, we say it is 'biased'. Bias is a systematic error (as opposed to a random error) that skews the observation to one side of the truth. Thus, if we use a scale that is not calibrated to zero, the weights we obtain using this scale will be biased. Similarly, if a sample is biased (for example, more males in the sample than the proportion of males in the population, or selecting cases from a hospital and controls from the general community in a case-control study), the results tend to be biased. Since it is often difficult to correct for biases once the data have been collected, it is always advisable to avoid bias when designing a study. (More details on biases and how to avoid them in Chapter 6.)

2.3 Experimental <u>versus</u> observational strategies

Although an experiment is an important step in establishing causality, it is often neither feasible nor ethical to subject human beings to risk factors in etiological studies. Instead, epidemiologists make use of 'natural experiments', when available, or they resort (more frequently) to analytical observational studies or quasi-experiments. However, there is one area of epidemiology in which experimental

strategies are used extensively: this is the area of clinical and field trials for testing new drugs or intervention programmes.

Advantages of the experimental approach include the following:

• The ability to manipulate or assign independent variables. This is by far the most distinct advantage of experimental strategies. It is readily illustrated by clinical trials, described in Chapter 4, in which cases of a specific disease are deliberately assigned (in random order, or by matching) to treatment and to control groups. For example, in an evaluation of the efficiency of intrauterine devices, women of a certain age and with certain other characteristics may be assigned at random or in matched pairs to physicians and nurses. A criterion for evaluation, such as the frequency of complications, is compared in the two groups. It may also be possible to manipulate the degree of exposure or the dose of the treatment.

• The ability to randomize subjects to experimental and control groups. Randomization makes it more likely that the distribution of some extraneous variables will be equalized between the two groups, although it is still necessary in the analysis to compare the distribution of these variables to ensure the validity of inferences drawn from the study. It is also possible in experiments (and also in some observational studies) to use matching in conjunction with randomization. In addition, randomization provides a basis for the calculation of appropriate probabilities of error in the inference.

• The ability to control confounding and eliminate sources of spurious association. Most of the other factors that interfere with the association under study can more easily be controlled in experiments (especially in animals) than in observational studies.

• The ability to ensure temporality. Determining which variables precede and which are the consequences of the intervention is more feasible in experimental studies than it is in some analytical studies, particularly those of the case-control and cohort designs.

• The ability to replicate findings. Experiments are often more replicable than observational studies. Replication satisfies the consistency requirement in causation. In practice, however, few clinical trials are exactly replicated.

All in all, the evidence for causal relationship is more compelling if it comes from a carefully executed experimental study, because selection factors that inadvertently bias observational studies can virtually be eliminated by the process of randomization. However, other sources of bias are not automatically controlled by randomization.

The limitations of the experimental approach are sometimes overlooked, as the impressive advantages of experiments have led some people to reject evidence for causation if it is not based on experiments. If we were limited to the experimental approach, however, we would have to abandon most of the evidence upon which significant advances in public health have been made. Experiments also have the following limitations:

• Lack of reality. In most human situations, it is impossible to randomize all risk factors except those under examination. Observational methods deal with more realistic situations.

• Difficulties in extrapolation. Results of experiments in animal models, which are rigorously controlled, cannot readily be extrapolated to human populations.

• Ethical problems. In human experimentation, people are either deliberately exposed to risk factors (in etiological studies) or treatment is deliberately withheld from cases (intervention trials). It is equally unethical to test the efficiency or side-effects of new treatments without critical evaluation in a small group of human subjects. (See also Chapter 10.)

• Difficulties in manipulating the independent variable. It is virtually impossible, for instance, to assign smoking habits at random to the experimental and control groups.

• Non-representativeness of samples. Many experiments are carried out on captive populations or volunteers, who are not necessarily representative of the population at large. Experiments in hospitals (where the experimental approach is most feasible and is frequently used) suffer from several sources of selection bias.

2.4   Descriptive studies

Definition

When an epidemiological study is not structured formally as an analytical or experimental study, i.e. when it is not aimed specifically to test a hypothesis, it is called a descriptive study, and belongs to the observational category of studies. The wealth of material obtained in most descriptive studies allows the generation of hypotheses, which can then be tested by analytical or experimental designs. A survey, for example a prevalence survey, could also be defined as a descriptive study, as it covers the elements of descriptive study.

Conduct of descriptive studies

Descriptive studies entail the collection, analysis and interpretation of data. Both qualitative and quantitative techniques may be used, including questionnaires, interviews, observations of participants, and service statistics, as well as documents describing communities, groups, situations, programmes and other individual or ecological units. The distinctive feature of this approach is that its primary concern is with description rather than with the testing of hypotheses or proving causality. The descriptive approach may, nevertheless, be integrated with or supplement methods that address these issues, and may add considerably to the information base.

Kinds of descriptive studies

Case series

This kind of study is based on reports of a series of cases of a specific condition, or a series of treated cases, with no specifically allocated control group. These represent the numerator of disease occurrence, and should not be used to estimate risks.

In an attempt to make such series more impressive, clinicians may calculate proportional distribution, which consists simply of percentages of the total number of cases that belong to a specific category of age, sex, ethnic group or other characteristic. These numbers are not rates, because the denominator still represents the cases and not the population at risk.

Community diagnosis or needs assessment

This kind of study entails collection of data on existing health problems, programmes, achievements, constraints, social stratification, leadership patterns, focal points of resistance or high prevalence, or groups at highest risk. Its purpose is to identify existing needs and to provide baseline data for the design of further studies or action.

Epidemiological description of disease occurrence

This common use of the descriptive approach entails the collection of data on the occurrence and distribution of disease in populations according to specific characteristics of individuals (e.g. age, sex, education, smoking habits, religion, occupation, social class, marital status, health status, personality), place (rural/urban, local, subnational, national, international) and time (epidemic, seasonal, cyclic, secular). A description may also be given of familial characteristics such as birth order, parity, family size, maternal age, birth interval or family type.

Descriptive cross-sectional studies or community (population) surveys

Cross-sectional studies entail the collection of data on, as the term implies, a cross-section of the population, which may comprise the whole population or a proportion (sample) of it. Many cross-sectional studies do not aim at testing a hypothesis about an association, and are thus descriptive. They provide a prevalence rate at a particular point in time (point prevalence) or over a period of time (period prevalence). The study population at risk is the denominator for these prevalence rates.

Included in this type of descriptive study are surveys in which the distribution of a disease, disability, pathological condition, immunological condition, nutritional status, fitness, or intelligence, etc., is assessed. This design may also be used in health systems research to describe 'prevalence' by certain characteristics – pattern of health service utilization and compliance – or in opinion surveys. A common procedure used in family planning and in other services is the KAP survey (survey of knowledge, attitudes and practice).

<u>Ecological descriptive studies</u>

When the unit of observation is an aggregate (e.g. family, clan or school) or an ecological unit (a village, town or country) the study becomes an ecological descriptive study.

As mentioned earlier, hypothesis testing is not generally an objective of the descriptive study. However, in some of the above studies (cross-sectional surveys, ecological studies) some hypothesis testing may be appropriate. Moreover, description of the data is also an integral part of the analytical study.

2.5   Analytical strategies in epidemiology

Observational studies, where establishing a relationship (association) between a 'risk factor' (etiological agent) and an outcome (disease) is the primary goal, are termed analytical. In this type of study, hypothesis testing is the primary tool of inference. The basic approach in analytical studies is to develop a specific, testable hypothesis, and to design the study to control any extraneous variables that could potentially confound the observed relationship between the studied factor and the disease. The approach varies according to the specific strategy used.

2.5.1   Case-control studies

The simplest and most commonly used analytical strategy in epidemiology involves the case-control study. It is designed primarily to establish the causes of diseases by investigating associations between exposure to a risk factor and the occurrence of disease. The design is relatively simple, except that it is backward-looking (retrospective) based on the exposure histories of cases and controls. With this type of study, one investigates an association by contrasting the exposure of a series of cases of the specified disease with the exposure pattern of carefully selected control groups free from that particular disease (Figure 2.1). Data are analysed to determine whether exposure was different for cases and for controls. The risk factor is something that happened or began in the past, presumably before disease onset, e.g. smoking, or a previous infection or medication. Information about the exposure is obtained by taking a history and/or from records. Occasionally, the suspected factor or attribute is a permanent one, such as blood group, which can be ascertained by clinical or laboratory investigation. A higher frequency of the attribute or risk factor among

FIGURE 2.1   DESIGN OF A CASE-CONTROL STUDY

```
                                    ┌──────────────────────────────┐
                                    │ Expose (with characteristic  │
                                    │ or risk factor)              │
                                    └──────────────────────────────┘
┌──────────────────────┐
│ Cases                │
│ (those with condition)│
└──────────────────────┘
                                    ┌──────────────────────────────┐
                                    │ Unexposed (without characteristic │
                                    │ or risk factor)              │
                                    └──────────────────────────────┘


                                    ┌──────────────────────────────┐
                                    │ Expose (with characteristic  │
                                    │ or risk factor)              │
                                    └──────────────────────────────┘
┌──────────────────────┐
│ Controls             │
│ (those without condition)│
└──────────────────────┘
                                    ┌──────────────────────────────┐
                                    │ Unexposed (without characteristic │
                                    │ or risk factor)              │
                                    └──────────────────────────────┘
```

Example

```
                                    ┌──────────────────────────────┐
                                    │ Chewers of tobacco           │
                                    └──────────────────────────────┘
┌──────────────────────┐
│ Cases of oral cancer │
└──────────────────────┘
                                    ┌──────────────────────────────┐
                                    │ Non-chewers of tobacco       │
                                    └──────────────────────────────┘


                                    ┌──────────────────────────────┐
                                    │ Chewers of tobacco           │
                                    └──────────────────────────────┘
┌──────────────────────┐
│ Those free of oral cancer│
└──────────────────────┘
                                    ┌──────────────────────────────┐
                                    │ Non-chewers of tobacco       │
                                    └──────────────────────────────┘
```

cases than among controls is indicative of its association with the disease/condition – an association that may be of etiological significance. In other words, if a greater proportion of cases than controls give a history of exposure, or have records or indications of exposure in the past, the factor or attribute can be suspected of being a causative factor.

Selection of cases

What constitutes a case in the study should be clearly defined with regard to the histological type and other specific characteristics of the disease, such as date of diagnosis, geographical location, etc. Cases that do not fit these criteria should be excluded from the study. This design is particularly efficient for rare diseases, because all cases that fit the study criteria in a particular setting within a specific period are usually included. This allows for a reasonable number of cases to be included in the study without waiting for the occurrence of new cases of the disease, which might take a long time.

For reasons of convenience and completeness of case records, the cases identified for case-control studies are often those from a hospital setting, from physicians' private practices, or from disease registries. Newly diagnosed cases within a specific period (incident cases) are preferred to prevalent cases, since such a choice may eliminate the possibility that long-term survivors of a disease were exposed to the investigated risk factor after the onset of the disease.

The selection of cases should be such that the study results are reliable and valid. For these reasons, the following guidelines should be used when selecting cases in a case-control study:

a. The criteria for inclusion in the study (what constitutes a case) and criteria for exclusion from the study must be clearly specified; this will improve the validity of the results;

b. The sources of cases may be:

- all cases admitted to or discharged from a hospital, clinic, or private practice within a specified period;

- all cases reported or diagnosed during a survey or surveillance programme within a specified period;

- incident or newly diagnosed cases;

- incident cases in an ongoing cohort study or in an occupational cohort (sometimes called a nested case-control study);

- deaths with a record of causes of death, and fulfilling other criteria for the study;

- case units with a prescribed health outcome;

c. If the number of cases is too large, a probability sample may be used;

d. Cases selected for the study should be representative of all cases of the disease under consideration.

Selection of controls

It is crucial to set up one or more control groups of people who do not have the specified disease or condition in order to obtain estimates of the frequency of the attribute or risk factor for comparison with its frequency among cases. This is the most important aspect of the case-control study, as biases in the selection of controls may invalidate the study results, and bias in the selection of controls is often the greatest cause for concern when analysing data from case-control studies.

a. The sources of comparison groups may be:

- a probability sample of a defined population, if the cases are drawn from that defined population;

- a sample of patients admitted to, or attending the same institution as the cases;

- a sample of relatives or associates of the cases (neighbourhood controls);

- a group of persons selected from the same source population as the cases, and matched with the cases for potentially confounding variables;

- on other risk factors (other than the one under consideration);

b. The selection of controls may involve matching on other risk factors:

- Matching means that controls are selected such that cases and controls have the same (or very similar) characteristics other than the disease and the risk factor being investigated. The characteristics are those that would confound the effect of the putative risk factor,

i.e. these characteristics are known to have an association with the disease, and may be associated with the risk factor being studied. The purpose of the matching is to ensure comparability of these characteristics for the two groups, so that any observed association between the putative risk factor and the disease is not affected by differential distribution of these other characteristics. It is common to match for age, sex, race and socioeconomic status in case-control studies on diseases, as we know all of these factors affect the incidence of most of the diseases. Matching may be done on an individual basis (one-to-one matching) or on a group basis (frequency matching). Individual matching is preferable, because of the ease of analysis accounting for matching. The disadvantages of matching include a loss of precision and overmatching. Also, once a matched design is used, the matching variable is eliminated from consideration, and therefore it cannot be investigated for etiological association with the disease. For example, if we matched for marital status in a study of breast cancer, we would not know whether single or married women had different risks for breast cancer. Many epidemiologists prefer to conduct studies without matching, and use statistical methods to adjust for possible confounding during analysis, because of the increased precision and the ability to investigate any possible interaction effects. The use of unmatched controls, obtained through random sampling, allows greater flexibility in studying various interactions. What is most important is that information on potential confounding factors should be collected in the study, so that these can be adjusted in the analysis.

c.   The number of control groups may vary. It is sometimes desirable to have more than one control group, representing a variety of disease conditions other than that under study and/or non-hospitalized groups. Use of multiple controls confers three advantages:

•   If the frequency of the attribute or risk factor does not differ from one control group to another, but is consistently lower than that among the cases, this increases the internal consistency of the association;

- If a control group is taken of patients with another disease, which is independently associated with the risk factor, the difference in the frequency of the factor between cases and controls may well be masked. In such a case, the use of another control group will save the research project;

- Multiple controls provide a check on bias.

The impact of poorly chosen controls on the conclusions of a case-control study is commonly exemplified by Pearl's study in 1929. Pearl compared 816 malignancies identified among 7500 autopsied cases at the Johns Hopkins Hospital in Baltimore, Maryland, USA with 816 non-malignant autopsied cases matched at death for age, sex, race and date of death. Lesions of active tuberculosis were found in 6.6% of cases and in 16.3% of controls, which led to the conclusion that there was antagonism between tuberculosis and cancer. This finding could not be corroborated in animal experiments. One explanation for Pearl's findings is that his control group inadvertently included many individuals who had died of tuberculosis, because tuberculosis patients were more frequently autopsied than were patients with other causes of death, and were thus unrepresentative of the general population of deaths.

Collection of data on exposure and other factors

Often data are collected through interviews, questionnaires and/or examination of records. Occasionally, clinical and laboratory examinations are carried out, but often this is not possible, especially if the 'cases' include past cases which may also include some deaths. The following precautions should be taken when deciding on the data-collection strategy:

- observation should be objective, or, if obtained by survey methods, well standardized;

- the investigator or interviewer should not know whether a subject is in the case or control group (blinding);

- the same procedures, e.g. interview and setting, should be used for all groups.

Multifactorial case-control studies

The common form of case-control study addresses one main factor or attribute at a time. It is possible, however, to investigate several exposure factors in the same study. For example, in a study in three states in the USA with a population of 13 million, all mothers of leukaemic children of 1-4 years old (diagnosed in 1959-67) were interviewed. As controls, a sample of 13 000 other women was taken. Four factors were considered, two preconceptional (preconceptional radiation and previous reproductive wastage) and two post-conceptional (in utero irradiation and viral infection during pregnancy). Analysis showed that each factor was related to leukaemia in their children (Gibson et al.,1968). Further analysis was conducted for combinations of factors, where the estimated relative risk in the absence of any of the four factors was made equal to 1.0, as shown in Table 2.1.

It is apparent that the effect was the greatest among women with all four factors, and that there is synergism between the factors.

Advantages of case-control studies

The following are examples of the advantages of case-control studies:

- feasible when the disease being studied occurs only rarely, e.g. cancer of a specific organ;

- relatively efficient, requiring a smaller sample than a cohort study;

- little problem with attrition, as when follow-up requires periodic investigations and some subjects refuse to continue to cooperate;

- sometimes they are the earliest practical observational strategy for determining an association (e.g. use of diethylstilbesterol and clear-cell adenocarcinoma of the vagina in daughters).

Enhancement of the validity of case-control studies

Ways in which one can increase the validity of a study include ensuring that:

- the cases are representative of all cases in a particular setting;

- the controls are similar to cases with respect to risk factors other than the study factor;

- multiple controls are used with consistent results;

- cases and controls are truly selected independently of exposure status;

- the sources of bias are mitigated, or at least shown not to have affected the results. (A common example is the British study of smoking and lung cancer by Doll and Hill (1952). After the cases and controls had been interviewed, it was discovered that some of the cases had been wrongly diagnosed as cancer. Reanalysis showed the persistence of the association and indicated that, in the study, the fact of being told that they had lung cancer did not bias the respondents with regard to the history they gave of smoking);

- repeated studies in different settings and by different investigators confirm each other (for example, the association between smoking and lung cancer has been reported by over 25 investigators from ten countries);

- it is possible to demonstrate a dose-response or gradient relationship (for example, several case-control studies showed that the number of cigarettes smoked per day was related to the risk of lung cancer);

- a hybrid design of case-control study nested in a cohort study with a defined population is used; this is a most powerful strategy.

Disadvantages and biases of case-control studies

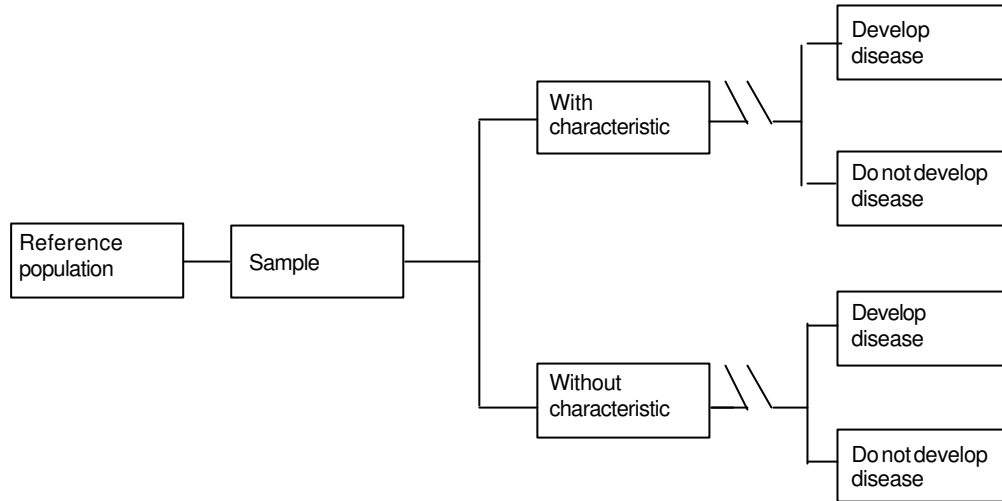The following are some of the problems associated with case-control studies:

- the absence of epidemiological denominators (population at risk) makes the calculation of incidence rates, and hence of attributable risks, impossible;

- temporality is a serious problem in many case-control studies where it is not possible to determine whether the attribute led to the disease/condition, or vice versa;

- there is a great risk of bias in the selection of cases and controls. This is particularly serious when a single control group is related to the risk factor under investigation;

- it may be very difficult or impossible to obtain information on exposure if the recall period is long;

- selective survival, which operates in case-control studies, may bias the comparison; there is no way of ascertaining whether the exposure was the same for those who died and those who survived;

- because most case-control studies are performed in hospitals, they are liable to Berkson's fallacy, or the effect of differing admission policies and rates;

- measurement bias may exist, including selective recall and misclassification (putting cases in the control group, or vice versa); there is also the possibility of the Hawthorne effect: with repeated interviews, respondents may be influenced by being under study;

- case-control studies are incapable of disclosing other conditions related to the risk factor: for example, in a study of the side-effects of oral contraceptives, one has to know their side-effects before a case-control design can be set up.

2.5.2 Prospective cohort studies

The common strategy of cohort studies is to start with a reference population (or a representative sample thereof), some of whom have certain characteristics or attributes relevant to the study (exposed group), with others who do not have those characteristics (unexposed group). Both groups should, at the outset of the study, be free from the condition or conditions under consideration. Both groups are then observed over a specified period to find out the risk each group has of developing the condition(s) of interest. This is illustrated diagrammatically in Figure 2.2.

FIGURE 2.2   DESIGN OF A COHORT (PROSPECTIVE) STUDY

```
                                    ┌─────────────┐
                                    │ Develop     │
                              ╲╱────│ disease     │
                 ┌──────────────┐   └─────────────┘
                 │ With         │
                 │ characteristic│  ┌─────────────┐
                 └──────────────┘   │ Do not develop│
                              ──────│ disease      │
                                    └─────────────┘
┌────────────┐  ┌──────────┐
│ Reference  │──│ Sample   │
│ population │  │          │        ┌─────────────┐
└────────────┘  └──────────┘        │ Develop     │
                              ╲╱────│ disease     │
                 ┌──────────────┐   └─────────────┘
                 │ Without      │
                 │ characteristic│  ┌─────────────┐
                 └──────────────┘   │ Do not develop│
                              ──────│ disease      │
                                    └─────────────┘
```

Example

```
                                    ┌─────────────┐
                                    │ Develop oral│
                              ╲╱────│ cancer      │
                 ┌──────────────┐   └─────────────┘
                 │ Chew tobacco │
                 └──────────────┘   ┌─────────────┐
                              ──────│ Do not develop│
                                    │ oral cancer  │
                                    └─────────────┘
┌────────────┐  ┌──────────┐
│ Population (free│ Sample   │
│ of oral  cancer│          │        ┌─────────────┐
└────────────┘  └──────────┘        │ Develop oral│
                              ╲╱────│ disease     │
                 ┌──────────────┐   └─────────────┘
                 │ Do not chew  │
                 │ tobacco      │   ┌─────────────┐
                 └──────────────┘   │ Do not develop│
                              ──────│ oral cancer  │
                                    └─────────────┘
```
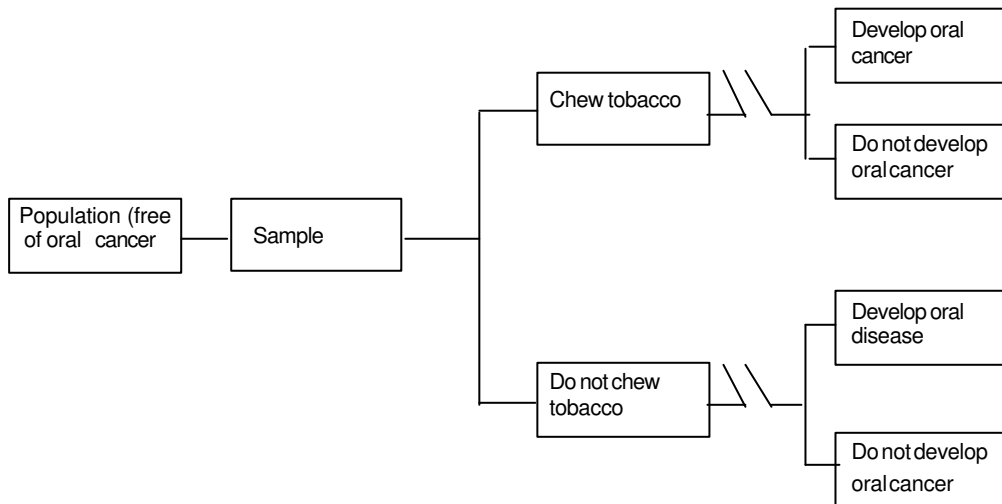
Design features

a.   Selection of cohort:

- a community cohort of specific age and sex;

- an exposure cohort, e.g. radiologists, smokers, users of oral contraceptives;

- a birth cohort, e.g. school entrants;

- an occupational cohort, e.g. miners, military personnel;

- a marriage cohort;

- a diagnosed or treated cohort, e.g. cases treated with radiotherapy, surgery, hormonal treatment.

The usual procedure is to locate or identify the cohort, which may be a total population in an area or sample thereof.

b.   Data to be collected:

- data on the exposure of interest to the study hypotheses;

- data on the outcome of interest to the study hypotheses;

- characteristics of the cohort that might confound the association under study.

c.   Methods of data collection

Several methods are used to obtain the above data, which should be on a longitudinal basis.   These methods include:

- interview surveys with follow-up procedures;

- medical records monitored over time;

- medical examinations and laboratory testing;

- record linkage of sets with exposure data and sets with outcome data, e.g. work history data in underground mines with mortality data from national mortality files.

In a conventional cohort study, an initial cross-sectional study is often performed to exclude persons with the outcome of interest (disease) and to identify the cohort that is free from the disease.

Measures of frequency

Two methods are commonly used in cohort studies to measure the incidence of the disease (condition) under investigation:

a. Cumulative incidence

    This index of disease frequency is based on the total population at risk which was, at entry to the study, free of the disease under investigation. The incidence of the disease is calculated for each stratum of exposure to the risk factor, and is the ratio of the number of new cases or events in a specified period of observation, to the total population at risk during that period.

    This incidence measure provides an estimate of the probability or risk of developing disease among all members of the group who were included in the study at its initiation, and were at risk of disease. Because cumulating all new cases in the total population at risk derives the measure, the term 'cumulative incidence' has been applied. Cumulative incidence is a proportion, not a rate, and can vary from 0 to 1, that is, no less than 0% and no more than 100% of the population at risk can acquire the disease.

    This measure of disease frequency is calculated as if all units or individuals had the same period of observation, but new cases are no longer at risk once they develop the disease.

b. Incidence density (person-time approach)

    This approach is an improvement over the conventional measure of incidence, because it takes into consideration both the number observed and the duration of observation for each individual. Thus, if 30 individuals were observed as follows: 10 for two years, 5 for three years, and 15 for four years, they would contribute $(10x2)+(5x3)+(15x4) = 95$ person-years of observation, which would become the denominator. The numerator is the number of new cases observed in these groups over the specified period of time. This gives an incidence rate per person-year, called the incidence density. Person-years do not represent the number of persons: 400 person-years of observation could represent 400 persons each observed for one year, or 40 persons each observed for 10 years. Two drawbacks of this measure are that the exact time when the disease occurs often cannot be ascertained, and that the rate of disease development over time is not necessarily constant.

The basic measures of effect used in cohort studies are the relative risk (RR), attributable risk (AR), population attributable risk (PAR), population attributable risk percent (ARP%), and etiologic fraction (EF). These measures will be discussed in detail in Chapter 7.

Advantages of cohort studies

The following are some of the advantages of a cohort study compared with a case-control study:

- Because of the presence of a defined population at risk, cohort studies allow the possibility of measuring directly the relative risk of developing the condition for those who have the characteristic, compared to those who do not, on the basis of incidence measures calculated for each of the groups separately.

- In a cohort study, it is known that the characteristic precedes the development of the disease, since all the subjects are free of disease at the beginning of the study; this allows for a conclusion of cause-effect relationship (a necessary, but not sufficient, condition).

- Because the presence or absence of the risk factor is recorded before the disease occurs, there is no chance of bias being introduced due to awareness of being sick as in encountered in case-control studies.

- There is also less chance of encountering the problem of selective survival or selective recall, although selection bias can still occur because some subjects who contracted the disease will have been eliminated from consideration at the start of the study.

- Cohort studies are capable of identifying other diseases that may be related to the same risk factor.

- Unlike case-control studies, cohort studies provide the possibility of estimating attributable risks, thus indicating the absolute magnitude of disease attributable to the risk factor.

- If a probability sample is taken from the reference population, it is possible to generalize from the sample to the reference population with a known degree of precision.

Disadvantages of cohort studies

The following are some of the disadvantages of cohort studies:

- These studies are long-term and are thus not always feasible; they are relatively inefficient for studying rare conditions.

- They are very costly in time, personnel, space and patient follow-up.

- Sample sizes required for cohort studies are extremely large, especially for infrequent conditions; it is usually difficult to find and manage samples of this size.

- The most serious problem is that of attrition, or loss of people from the sample or control during the course of the study as a result of migration or refusal to continue to participate in the study. Such attrition can affect the validity of the conclusion, if it renders the samples less representative, or if the people who become unavailable are different from those actually followed up. The higher the proportion lost (say beyond 10-15%) the more serious the potential bias.

- There may also be attrition among investigators who may lose interest, leave for another job, or become involved in another project.

- Over a long period, many changes may occur in the environment, among individuals or in the type of intervention, and these may confuse the issue of association and attributable risk.

- Over a long period, study procedures may influence the behaviour of the persons investigated in such a way that the development of the disease may be influenced accordingly (Hawthorne effect). This problem is more likely to occur in studies involving repeated contact with participants, as in studies of diet or the use of contraceptives. The participants may modify their diet or shift to another contraceptive method because of repeated probing. Behavioural changes are also a serious problem in opinion surveys, acceptability studies and psychological investigations, such as studies of the psychological sequelae of sterilization.

- A serious ethical problem may arise when it becomes apparent that the exposed population is manifesting significant disease excess before the follow-up period is completed.

It must be emphasized that, although the cohort study is close to the randomized trial (experiment) in terms of epidemiological power, it may still have problems of validity. Care must be taken to ensure that it satisfies other requirements of epidemiological research, particularly with regard to appropriate sampling, construction of comparison groups, handling of missing data, application of appropriate statistical methods and other prerequisites for a sound research design.

2.5.3 Historical (retrospective) cohort studies

In a prospective cohort study, the investigators or their substitutes are typically present from the beginning to the end of the observation period. However, it is possible to maintain the advantages of the cohort study without the continuous presence of the investigators, or having to wait a long time to collect the necessary data, through the use of a historical or retrospective cohort study. The design of such a study is illustrated in Figure 2.3.
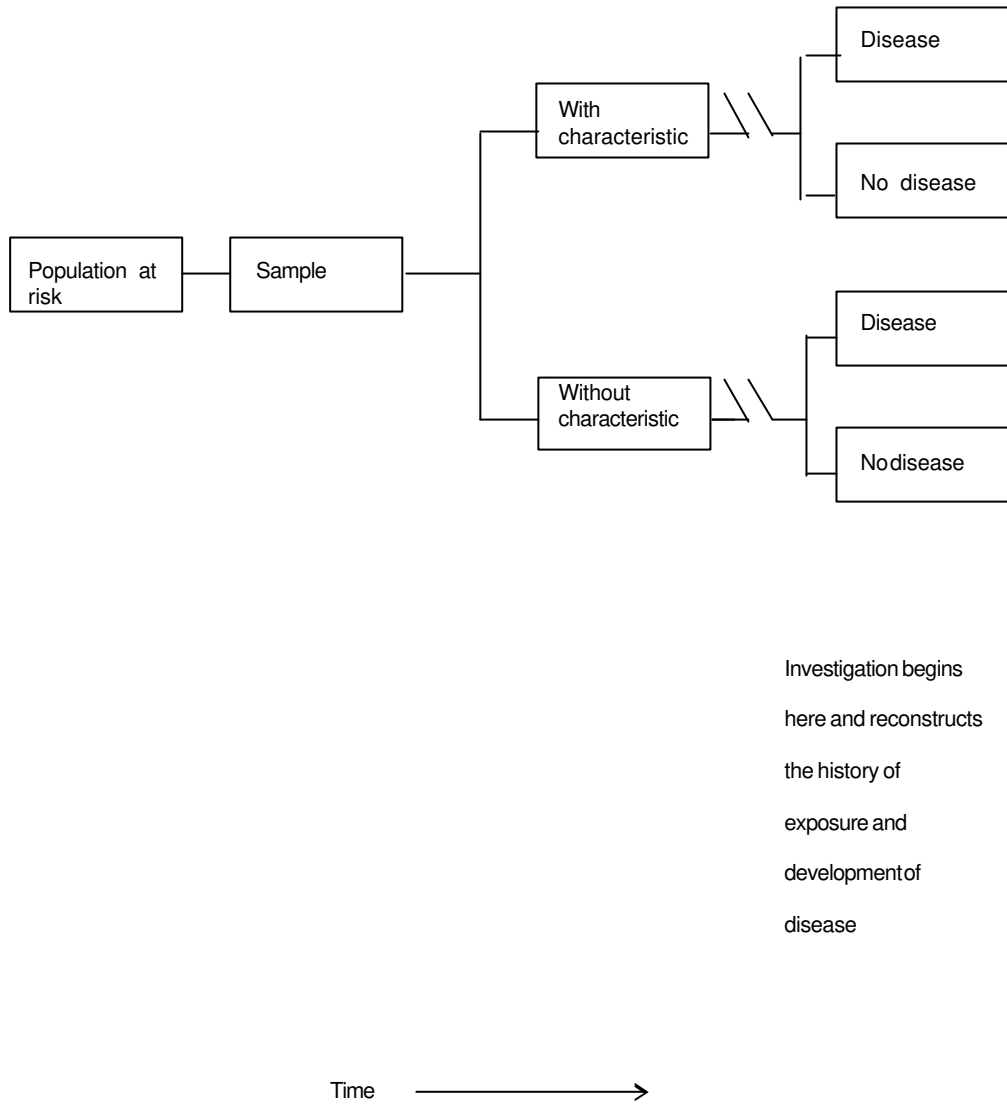
A historical cohort study depends upon the availability of data or records that allow reconstruction of the exposure of cohorts to a suspected risk factor and follow-up of their mortality or morbidity over time. In other words, although the investigator was not present when the exposure was first identified, he reconstructs exposed and unexposed populations from records, and then proceeds as though he had been present throughout the study.

Historically constructed cohorts share several advantages of the prospective cohort. If all requirements are satisfied, a historical cohort may suffer less from the disadvantages of time and expense. Historical cohort studies have, however, the following disadvantages:

- All of the relevant variables may not be available in the original records.

- It may be difficult to ascertain that the study population was free from the condition at the start of the comparison. This problem does not exist if we are concerned with deaths as indicators of disease.

FIGURE 2.3   DESIGN OF A HISTORICAL (RETROSPECTIVE) COHORT STUDY

```
                                                          ┌──────────────┐
                                                          │   Disease    │
                                          ┌────────────┐──┤              │
                                          │    With    │  └──────────────┘
                                    ┌─────┤characteristic├──\\ 
                                    │     └────────────┘  ┌──────────────┐
                                    │                     │  No disease  │
┌──────────────┐   ┌──────────┐     │                     └──────────────┘
│ Population at │───│  Sample  │─────┤
│ risk          │   └──────────┘     │                     ┌──────────────┐
└──────────────┘                     │                     │   Disease    │
                                     │     ┌────────────┐──┤              │
                                     └─────┤  Without   ├──\\ └──────────────┘
                                           │characteristic│  ┌──────────────┐
                                           └────────────┘  │  No disease  │
                                                            └──────────────┘
```

Investigation begins

here and reconstructs

the history of

exposure and

development of

disease

Time   ⟶

- Attrition problems may be serious due to loss of records, incomplete records, or difficulties in tracing or locating all of the original population for further study.

- These studies require ingenuity in identifying suitable populations and in obtaining reliable information concerning exposure and other relevant factors. Examples of such population groups include members of health insurance plans, military personnel, industrial groups (such as miners), professional groups, members of a trade union, etc.

2.5.4 Prognostic cohort studies

Prognostic cohort studies are a special type of cohort study used to identify factors that might influence the prognosis after a diagnosis or treatment. These follow-up studies have the following features:

- The cohort consists of cases diagnosed at a fixed time, or cases treated at a fixed time by a medical or surgical treatment, rehabilitation procedure, psychological adjustment or vocational adjustment.

- By definition, such cases are not free of a specified disease, as in the case of a conventional cohort study (but are free of the 'outcome of interest').

- The outcome of interest is usually survival, cure, improvement, disability, vocational adjustment, or repeat episode of the illness, etc.

2.5.5 Analytical cross-sectional studies

In an analytical cross-sectional study, the investigator measures exposure and disease simultaneously in a representative sample of the population. By taking a representative sample, it is possible to generalize the results obtained in the sample for the population as a whole. Cross-sectional studies measure the association between the exposure variable and existing disease (prevalence), unlike cohort studies, which measure the rate of developing disease (incidence). Rare diseases, conditions of short duration, or diseases with high case fatality are often not detected by the one-time snapshot of the cross-

sectional study. Therefore, cross-sectional studies are more appropriate for measuring the relationship between fairly permanent characteristics in individuals and chronic diseases or stable conditions.

Design

Cross-sectional studies are represented in Figure 2.4. They usually start with a reference population, from which a random sample is taken. Data are collected at the same time on the risk factor or characteristic and the condition.

FIGURE 2.4   DESIGN OF A CROSS-SECTIONAL STUDY
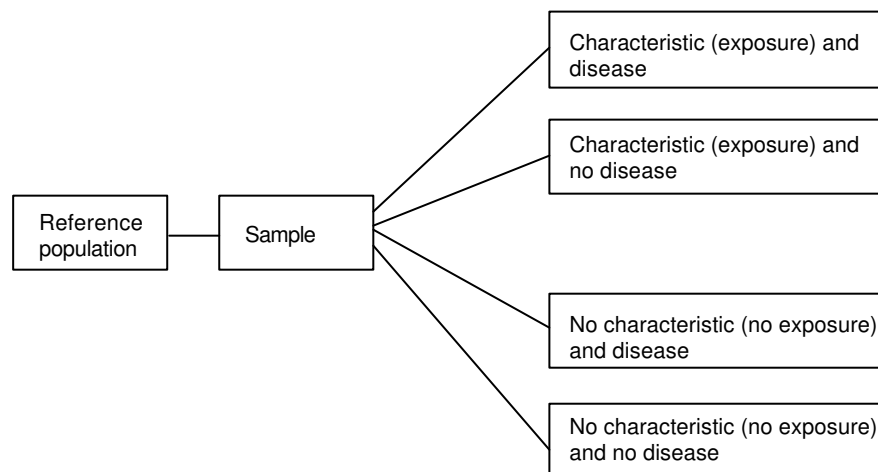
Advantages of cross-sectional studies

The following are some advantages of cross-sectional studies:

- Cross-sectional studies have the great advantage over case-control studies of starting with a reference population from which the cases and controls are drawn.

- They can be short-term, and therefore less costly than prospective studies.

- They are the starting point in prospective cohort studies for screening out already existing conditions.

- They provide a wealth of data that can be of great use in health systems research.

- They allow a risk statement to be made, although this is not precise.

Disadvantages of cross-sectional studies:

- They provide no direct estimate of risk.

- They are prone to bias from selective survival.

- Since exposure and disease are measured at the same point in time, it is not possible to establish temporality (i.e. whether the exposure or presence of a characteristic preceded the development of the disease or condition).

2.5.6 Ecological studies

In ecological studies, the unit of observation is an aggregate, a geographical administrative locality, a cluster of houses, a town, a whole country, etc. They may take any of the following forms:

- descriptive

- case-control

- cross-sectional

- cohort, or

- experimental.

Some specific forms of ecological studies are discussed below.

Aggregate analysis of national figures

These studies consist of an aggregate analysis of the correlation between a study factor and a disease (or mortality from a specific cause) in the geographical locale. They do not offer information on the exposure status of the individuals afflicted with or dead from the specific cause. Instead, the level of experience in the geographical unit or country is taken as a surrogate measure for all the individuals in that unit or country. Examples include:

- ecological correlation of per capita consumption of cigarettes and level of mortality from lung cancer;

- ecological correlation of water hardness and mortality from cardiovascular disease;

- maps of cancer frequency in a country and their interpretation by national cancer research authorities;

- ecological correlation of birth rate with gainful employment of women outside the home.

Time-series ecological studies

A variety of ecological studies may add a time-series dimension by examining, still on an aggregate basis, whether the introduction of a factor into a geographical area was associated with an increase in morbidity or mortality, or whether intervention in a geographical area reduced the morbidity or mortality. A good example is the study of death certificates for US women of reproductive age between 1961 and 1966 (Markush and Siegel, 1969), to find out whether there had been an increase in mortality from thromboembolism in women after the introduction of oral contraceptives in 1960-61.

Disadvantages and biases in ecological studies

While such studies are of interest as sources of hypotheses and as initial or quick methods of examining associations, they cannot be used as the basis for making causal inference. Their most serious flaw is the risk of ecological fallacy, when the characteristics of the geographical unit are incorrectly attributed to the individuals. Other sources of confounding are possible since many risk factors have a

tendency to cluster in certain geographic areas. Thus, air pollution, heavy industry, ageing and crowding correlate to cities. The death of a person from heart disease may have little or no relationship to the presence of heavy industry.

2.6 Comparison of the three major analytical strategies

The major attributes of the three major strategies, the case-control, cohort and cross-sectional study, are outlined in Table 2.2. Note that an experiment (a clinical trial, for example) has the same properties as the prospective cohort study, except that the exposure variable (usually an intervention) is deliberately assigned to experimental and control groups.

**TABLE 2.1 ESTIMATED RELATIVE RISKS FOR LEUKAEMIA IN CHILDREN
1-4 YEARS OF AGE FOR COMBINATIONS OF RISK FACTORS**

No. of preconceptional factors	No. of post-conceptional factors		
	None	**One**	**Two**
None	1.0	1.1	1.8
One	1.2	1.6	2.7
Two	1.9	3.1	4.6

TABLE 2.2 COMPARISON OF THREE ANALYTICAL STRATEGIES

Attribute	Type of analytical strategy		
	Cohort	**Case-control**	**Cross-sectional**
Classification of population	Population free from condition or disease, with or without characteristic	Cases with condition (disease) with or without the characteristic, and controls	Populations without identification of condition or characteristic
Sample represented	Non-diseased	Uncertain: the source population of the cases is unknown	Survivors at a point or period in time
Temporal sequence	Prospective or retrospective	Retrospective	Contemporary or retrospective
Function	Compares incidence rates in exposed and unexposed	Compares prevalence of exposure among cases and controls	Describes association between exposure and disease simultaneously
Outcome	Incidence of disease in exposed and unexposed	Prevalence of exposure in cases and controls	Prevalence of disease in exposed and unexposed
Risk measure	Relative risk, attributable risk	Odds ratio (estimate of relative risk)	Prevalence ratio (inexact estimate of relative risk); also odds ratio
Evidence of causality	Strong	Needs more careful analysis	Only suggestive
Bias	Easy to manage	Needs more effort and sometimes very difficult to manage	May be very difficult to manage

2.7 Choice of strategy

The bases for choosing one of the research strategies are summarized in Table 2.3.

TABLE 2.3   CHOICE OF STRATEGY

Basis	Cohort	Case-control	Cross-sectional
Rare condition	Not practical	Bias	Not appropriate
To determine a precise risk	Best	Only estimate possible	Gives relative prevalence, not incidence
To determine whether exposure preceded disease	Best	Not appropriate	Not appropriate
For administrative purposes	Not appropriate	Not appropriate	Best
If attrition is a serious problem	Not appropriate	Attrition is usually minimal	Attrition may have occurred before the study
If selective survival is problem	Best	Not appropriate	Not appropriate
If all factors are not known	Best	Not appropriate	Less appropriate
Time and money	Most expensive	Least expensive	In between

2.8   References and further reading

Doll R., Hill A.B. A study of the aetiology of carcinoma of the lung. British Medical Journal, 1952, 2.

Gibson R.W. et al. Leukemia in children exposed to multiple risk factors. New England Journal of Medicine, 1968, 279: 906-909.

Kleinbaum D.G., Kupper L.L., Morgenstern H. Epidemiologic research: principles and quantitative methods. London, Lifetime Learning Publications, 1982.

Markush R., Siegel D. Oral contraceptives and mortality trends from thromboembolism in the United States. American Journal of Public Health, 1969, 59: 418-434.

Schlesselman J.J. Case-control studies. New York, Oxford University Press, 1982.

Chapter 3

Descriptive Epidemiological Studies

3.1 Introduction

As mentioned in Chapter 2, a descriptive epidemiological study is usually a precursor to the analytical study testing hypotheses. In descriptive studies, morbidity or mortality in the population is examined, and its occurrence and distribution in population groups according to (1) characteristics of persons, (2) characteristics of place, and (3) characteristics of time, are illustrated.

The numbers of events (mortality or morbidity) are enumerated and the population at risk identified. Rates, ratios and proportions are calculated as measures of the probability of events. One must be careful to use the right measurements and the right 'denominators' when assessing these measures of probability. Comparison of the measures of probability across subgroups of populations is performed to identify the variables (time, place and person) that might explain the variability in mortality and morbidity within and between population groups. In this chapter, the major issues involved in dealing with descriptive studies are discussed.

3.2 Types of descriptive studies

Case series

This kind of study is based on reports of a series of cases of a specific condition, or a series of treated cases, with no specifically allocated control group. They represent the numerator of disease occurrence, and should not be used to estimate risks.

The distribution of cases by important factors of time, place and person might produce initial suspicion regarding potential causes, and might lead to more detailed descriptive studies, from which hypotheses may be generated. This will then lead to a formal analytical study to test these hypotheses. For example, the initial observation on AIDS was from a case series in San Francisco; the distribution of cases was almost entirely among homosexual men, which led to the suspicion about sexual practices as a potential cause. When a series of cholera cases is reported from a particular area in a country, initial tabulation of the case series might lead to a potential source of the epidemic, and subsequent analytical studies would confirm or dispel the initial suspicions.

Community diagnosis or needs assessment

This kind of study entails collection of data on existing health problems, programmes, achievements, constraints, social stratification, leadership patterns, focal points of resistance or high prevalence, or groups at highest risk. Its purpose is to identify existing needs and to provide baseline data for the design of further studies or action.

A description of common problems in a specific subgroup of the population (e.g. the homeless) and the facilities available to help these people, might lead to community action to increase the awareness of the problem and mobilization of community resources to solve the problem.

Epidemiological description of disease occurrence

This common use of the descriptive approach entails the collection of data on the occurrence and distribution of disease in populations according to specific characteristics of individuals (e.g. age, sex, education, smoking habits, religion, occupation, social class, marital status, health status, personality), place (rural/urban, local, subnational, national, international) and time (epidemic, seasonal, cyclic, secular). A description may also be given of familial characteristics such as birth order, parity, family size, maternal age, birth interval or family type.

This is the most common use of descriptive epidemiological studies. The measures of disease occurrence, for example incidence and prevalence or mortality, are commonly reported from many jurisdictions. Careful analysis of these descriptive statistics would lead

to the formulation of hypotheses and testing of these hypotheses with analytical studies. Care should be taken as to what indices are used in determining the 'risks'. These will be discussed later in the chapter.

Descriptive cross-sectional studies or community (population) surveys

Cross-sectional studies entail the collection of data on, as the term implies, a cross-section of the population, which may comprise the whole population or a proportion (a sample). Many cross-sectional studies do not aim at testing a hypothesis about an association, and are thus descriptive. They provide a prevalence rate at a point in time (point prevalence) or over a period of time (period prevalence). The study population at risk is the denominator for these prevalence rates.

Included in this type of descriptive study are surveys, in which the distribution of a disease, disability, pathological condition, immunological condition, nutritional status, fitness, intelligence, etc., is assessed. This design may also be used in health systems research to describe 'prevalence' by certain characteristics – pattern of health service utilization and compliance – or in opinion surveys. A common procedure used in family planning and in other services, is the KAP survey (survey of knowledge, attitudes and practice).

Ecological descriptive studies

When the unit of observation is an aggregate (e.g. family, clan or school) or an ecological unit (a village, town or country) the study becomes an ecological descriptive study.

As mentioned earlier, hypothesis testing is not generally an objective of the descriptive study. However, in some of the above studies (cross-sectional surveys, ecological studies) some hypothesis testing may be appropriate. Moreover, description of the data is an integral part of the analytical study.

3.3   Measures of incidence and prevalence

These measures of the distribution of the occurrence of disease are probably the most commonly used indicator of morbidity in the population. Incidence measures the occurrence of new cases of a disease, and prevalence measures the existing cases of the disease.

3.3.1 Incidence from longitudinal studies

Incidence is a measure of the frequency with which new disease events occur, and the rate at which people free from the disease develop the disease during a specified period of observation. A period of one year is commonly used. The important aspects of this measure are:

- the need to define the population of interest; this is often called the inception cohort;

- all the persons in the inception cohort should be free of the disease;

- a period of observation should be specified;

- all persons should be followed for the specified period of observation;

- if incomplete follow-up is encountered (some followed up for less than the specified period), the estimates of the incidence rates should be appropriately adjusted (i.e. incidence density rather than cumulative density should be used).

Two common measures of incidence are used in descriptive studies: the cumulative incidence and the incidence density. When all the people in the population of interest have been followed up for the specified period, the number of new cases divided by the size of the population provides the cumulative incidence. This is a proportion and is a measure of risk of acquiring the disease in that population over the specified period.

If there are different periods of follow-up for different people, the denominator in the above calculation is adjusted as person-time (e.g. if 100 people are followed for 6 months, and 100 people are followed for one year, the total observation is 1800 person-months or 150 person-years). The resulting index is called the incidence density, and gives an estimate of the 'instantaneous probability' of acquiring the disease in that population.

Tabulation of incidence rates by various categories of person, place and time will be useful to identify potential causes (risk factors) in the variation of incidence, which might be used in subsequent studies to verify or establish the results.

3.3.2 Use of incidence rates for surveillance

Conventionally, incidence rates are used by health agencies for surveillance purposes. Annual incidence rates are computed and charted, and the variations in the annual incidence rates are used to identify potential problem areas by analysing the trends. For example, if the annual incidence rate for tuberculosis has been steady for some time, and suddenly an increase is noticed in a particular year, studies may be undertaken to identify the causes, and preventive actions instituted.

In certain recurrent events, such as the common cold, allergy or asthma, the number of 'episodes', rather than the number of 'cases' may be used in the numerator. Sometimes the term 'attack rate' is used for such rates. (See J. Last: Dictionary of Epidemiology for the various uses of these terms.)

Changes in incidence may occur with the following factors:

- introduction of a new risk factor (e.g. oral contraceptives and increase in thromboembolism; food additives and cancer);

- changing habits (e.g. increased smoking and lung cancer; fluoridated water and decrease in dental caries);

- changing virulence of causative organisms (e.g. drug-resistant bacteria and deaths from infection; drug resistance to malaria prophylaxis and increase in malaria);

- changing potency of treatment or intervention programmes (e.g. vaccination against measles decreased the incidence of measles; relaxation of anti-venereal disease campaigns and an increase in the incidence of VD);

- selective migration of susceptible persons to an endemic area.

3.4   Prevalence

Prevalence is a measure of the <u>status quo</u> of a disease in a population at a fixed point of time, or during a specified period. It is the proportion of people who have the disease at the specified point or period. Prevalence is valuable for administrative purposes, for example, for determining the workload of personnel in a health programme. It is also useful in 'community diagnosis', i.e. to identify communities that need special programmes or action to prevent general illness.

Prevalence rates are typically obtained from cross-sectional studies such as national health surveys. Occasionally, they are based on disease registries (national or population-specific). Prevalence depends on previous incidence (I) and the duration of the disease (D). When both the incidence and duration are relatively stable, P = I x D.

Prevalence may change over time, depending on:

- changes in incidence;

- changes in disease duration and chronicity (e.g. some diseases may become shorter in duration or more acute because of a high recovery rate or high case fatality rate);

- intervention programmes;

- selective attrition (e.g. selective migration of cases, or of susceptible or immune persons);

- changing classifications (this is particularly important when using routinely collected national statistics to monitor trends in prevalence; the data coding according to various disease categories often changes, and variations in prevalence may be reported due to misclassification).

3.5 Examples

The following example illustrate the differences between incidence and prevalence, and the calculation of incidence and prevalence rates in simple situations:

Example 1

Population, 1 January: 100

	1 January	1 July	31 December
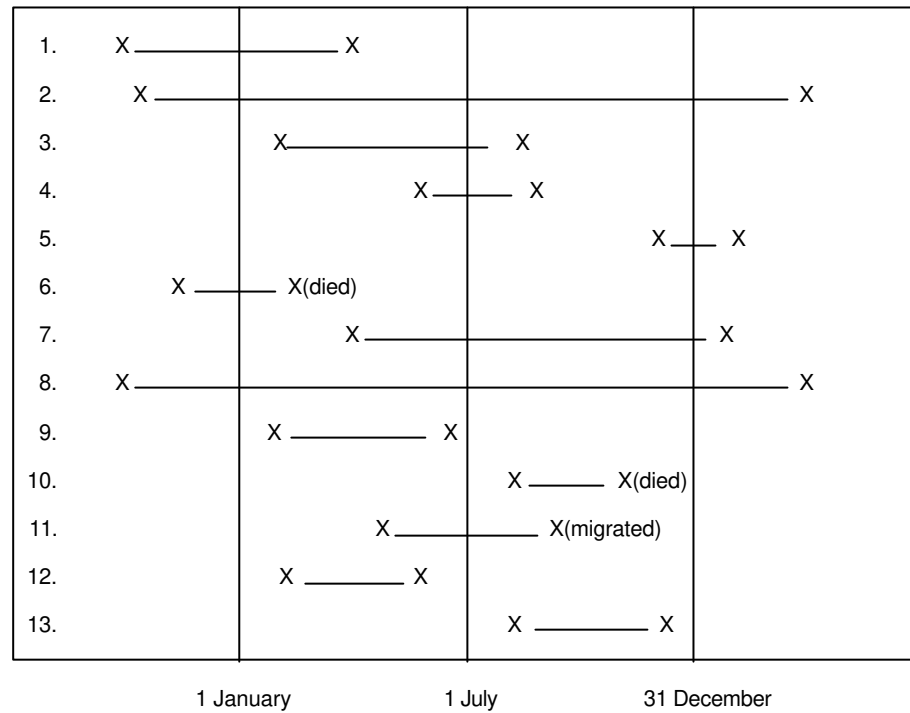
Point prevalence, 1 Jan = all cases per total population = 4 per 100 = 4%

Point prevalence, 1 Jul = all cases on 1 Jul per population on 1 Jul

= 5 / (100 - 2) = 5.1%

Point prevalence, 31 Dec = all cases on 31 Dec per population on 31 Dec

= 4 / (100 - 4) = 4.2%

Period prevalence in year = all cases in year per mid-year population

= (4 + 11) / (100 - 2) = 15.3%

Cumulative incidence = new cases during year / persons free from disease: 1 Jan

= 11 / (100 - 4) = 11.5%

<u>Example 2</u>

A population of 1000 females aged 40 years or over was screened for diabetes on 1 January 1998, and 40 cases were detected.   During the latter half of the year, five patients died, five migrated and five recovered. Meanwhile, 20 new cases were detected.   We want to measure the morbidity from diabetes in this group during 1998.   The flow chart shown in Figure 3.1 is a chronicle of the progression of events.

FIGURE 3.1 RESULTS OF SCREENING FOR DIABETES ON INCIDENCE RATE
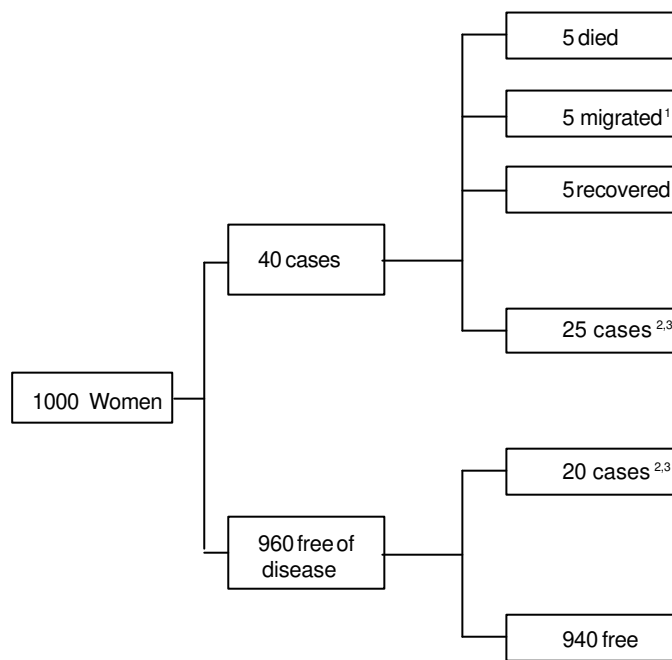
Screening 1 Jan 1998   31 December 1998

[1]Attrition

[2]Prevalent cases 31 December 1998

[3]Incident cases during 1998

Point prevalence on 1 January 1998   = 40 per 1000

Point prevalence on 31 Dec 1998   = (25 + 20) / 990   = 45.4 per 1000

Period prevalence 1998   = (40 + 20) / 1000 = 60 per 1000,

assuming all attrition occurred after mid-year.

Cumulative incidence during 1998   = 20 / 960 = 20.8 per 1000

Example 3

Divergence of incidence and prevalence trends

Suppose the results in Table 3.2 and Figure 3.3 were available for a childhood disease between 1983 and 1992.

TABLE 3.2   INCIDENCE AND PREVALENCE OF CHILDHOOD DISEASE X

Year	Incidence / 100 000	Prevalence / 100 000
1983	24.5	42.8
1984	24.9	41.2
1985	23.8	40.9
1986	24.6	40.1
1987	24.1	38.4
1988	24.7	37.9
1989	24.2	35.3
1990	23.9	33.2
1991	25.1	29.8
1992	24.5	27.2

FIGURE 3.3   INCIDENCE AND PREVALENCE OF CHILDHOOD DISEASE X 1983-1992
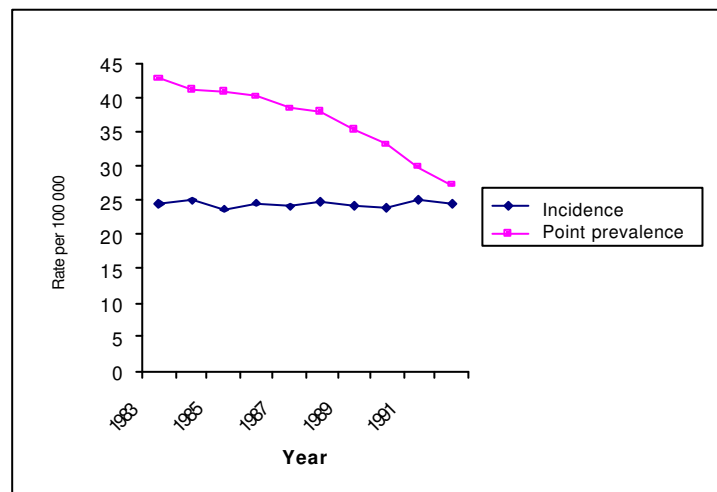
Interpretation

1. Recovery from the disease is becoming more rapid: for example, a new drug has been discovered that is being used more frequently.

2. The opposite situation is occurring: the disease is becoming more fatal (i.e. the case fatality
   ratio is increasing); for example, an increase in disease virulence, increasing failure of treatment, or decreasing application of effective treatment.

3. There is increasing, selective migration of cases (perhaps seeking treatment elsewhere).

Example 4

A disease in which the incidence over time is stable, while the prevalence is increasing, can be represented diagrammatically as shown in Figure 3.4:

**FIGURE 3.4   DISEASE IN WHICH INCIDENCE IS STABLE
AND PREVALENCE IS INCREASING**
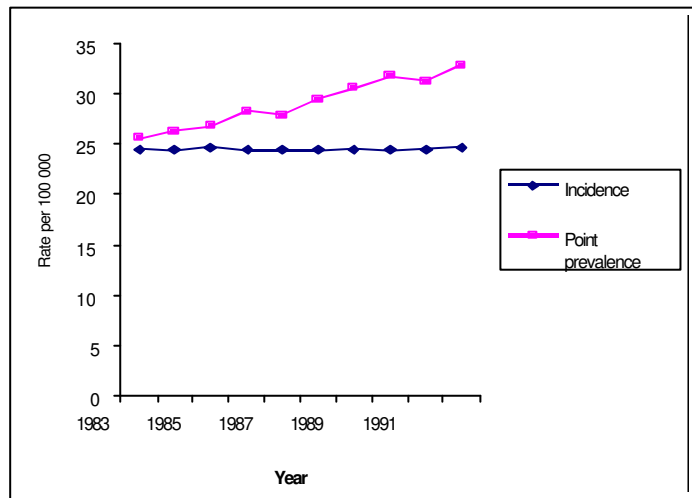
Interpretation

1. Recovery from the disease is becoming slower (i.e. the disease is becoming more chronic). For example, the drugs used are becoming less effective or are less frequently used, or resistance to the drugs is increasing.

2. The disease is becoming less fatal due, for example, to increased use of existing treatment, use of a newly discovered, potent drug that can affect the course but not the onset of the disease, or the organism is becoming less virulent.

3. There is selective immigration of cases from outside the area.

Example 5

FIGURE 3.5   DISEASES IN WHICH INCIDENCE IS INCREASING
AND PREVALENCE IS DECREASING
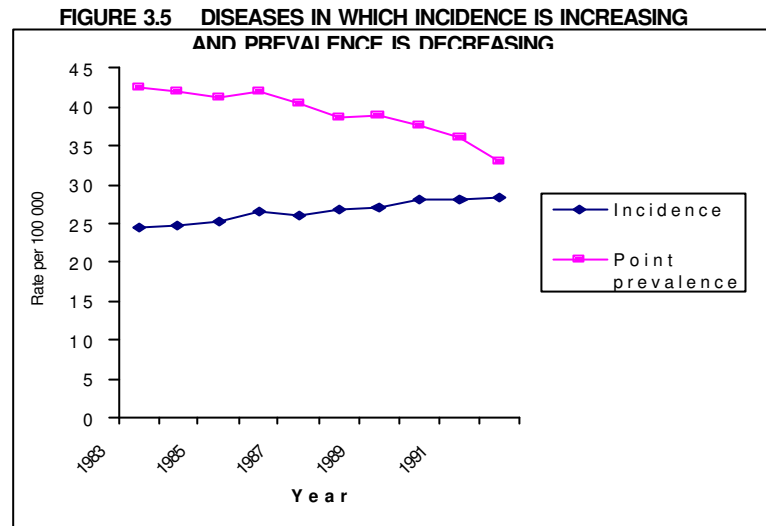
A case in which the incidence is increasing over time, but the prevalence is decreasing can be presented as shown in Figure 3.5.

Interpretation

1. The disease is becoming significantly shorter in duration: thus, while occurring more frequently, it is becoming more acute.

2. The disease is becoming more fatal.

3.6   **Comparison of rates**

It should be noted that in the above examples, crude rates were compared between years. This can be quite misleading, especially if the population structure has changed over the years. In epidemiology, when comparing rates between places or between times, it is important to take into account any concomitant changes in other related variables, primarily age, sex and race. This is commonly done by the

'standardization of rates' or by the use of multivariate mathematical models; this will be discussed later.

As an example, let us consider age. Age structure can affect incidence, prevalence and mortality. Hence, when comparing communities at a point in time, or the same community at different points in time, especially when age structure is variable, certain refinements in the measures of morbidity and mortality are necessary. These include:

- restriction of case comparison to one age group (e.g. comparing fertility of women aged 20-24, or blood pressure in males aged 50-59);

- use of age-specific rates;

- age adjustment of rates, using direct or indirect method of adjustment (standardization);

- matching for age at the stage of design; this will prevent the examination of age effects; and

- use of stratification analysis, or other multivariate analysis, in which age is one of the independent variables considered.

3.7   References and further reading

Hennekens C.H., Buring J. Epidemiology in medicine. Boston, Little, Brown and Company, 1987.

Beaglehole R., Bonita R., Kjellstrom T. Basic epidemiology. Geneva, WHO, 1993.

Greenberg R.S., Daniels S.R., Flanders W.D., Eley J.W., Boring J.R. Medical epidemiology, 2 ed, Norwalk, Appleton and Lange, 1996.

Chapter 4

Experimental and Quasi-experimental Studies

4.1 Introduction

As discussed in Chapter 2, an experiment is the best epidemiological study design to prove causation. It can be viewed as the final or definitive step in the research process, a mechanism for confirming or rejecting the validity of ideas, assumptions, postulates and hypotheses about the behaviour of objects, or effects upon them which result from interventions under defined sets of conditions. The experimenter (investigator) has control of the subjects, the intervention, outcome measurements, and sets the conditions under which the experiment is conducted. In particular, the investigator determines who will be exposed to the intervention and who will not. This selection is done in such a way that the comparison of outcome measure between the exposed and unexposed groups is as free of bias as possible.

As in other research designs, the investigator is rarely able to study all units within a population; a sample must be drawn from a target population for the purposes of the experiment, which will preserve the integrity of representativeness for the purposes of generalization. This is often done through a probabilistic process of random selection of study units. While this process cannot guarantee representativeness in small samples, in the long run the samples are expected to be unbiased and representative of the populations being studied.

In health research, we are often interested in comparative experiment, where one or more groups with specific interventions is compared with a group unexposed to interventions (placebo in clinical trials) or exposed to the best treatment currently available. The effect

of the new interventions on one or more outcome variables is compared between the groups by the use of statistical procedures, and the significance of observed differences assessed for concordance with the null hypotheses. Two types of comparative experiments, the randomized clinical trial (RCT) and the community intervention trial (CIT) are discussed in this chapter.

4.2   Purpose of experiment

The design of experiments serves the purpose of ensuring valid data relevant to the hypothesis under test as economically (maximum statistical power with minimum cost and inconvenience) as possible. A population survey tells us about the characteristics observed and the association between these characteristics in the population. But the value of this association is still a description without a causal meaning, e.g. that variable Y can be made to have a specific value by altering variable X. An experiment can tell us if we can make such causal associations.

The essence of the comparative experiment lies in the fact that we ensure that:

-       the only difference between the two groups is that of the intervention;

-       there is a sufficiently large number of units from the comparative groups;

-       appropriate probabilistic methods are employed to identify the relationship of the intervention to the outcome.

Careful choice of the outcome variable, the sample selection, allocation process and the statistical analysis procedures is essential for the success of the experiment. It is often necessary to deviate from this ideal condition of experiment, yet by careful choice of the design, it is often possible to adjust, statistically, any baseline differences or other anomalies in the sample selection process when testing the hypotheses.

Although an experiment is an important step in establishing causality, it must be remembered that it is often neither feasible nor ethical to subject human beings to risk factors in etiological studies. Experiments are therefore often confined to the clinical trial of new drugs, or therapies which would be potentially more beneficial, and

therefore ethically justifiable. Similarly, community interventions, such as fluoridation of water supply, introduction of educational programmes to instruct mothers in matters of nutrition, etc. are feasible experiments.

We will now discuss some general principles of experimentation and look at the RCT and CIT as specific examples.

4.3   The experimental design

The choice of a good experimental design depends on several factors. Figure 4.1 gives a flow chart of a general experimental design.

FIGURE 4.1   FLOW CHART OF AN EXPERIMENT
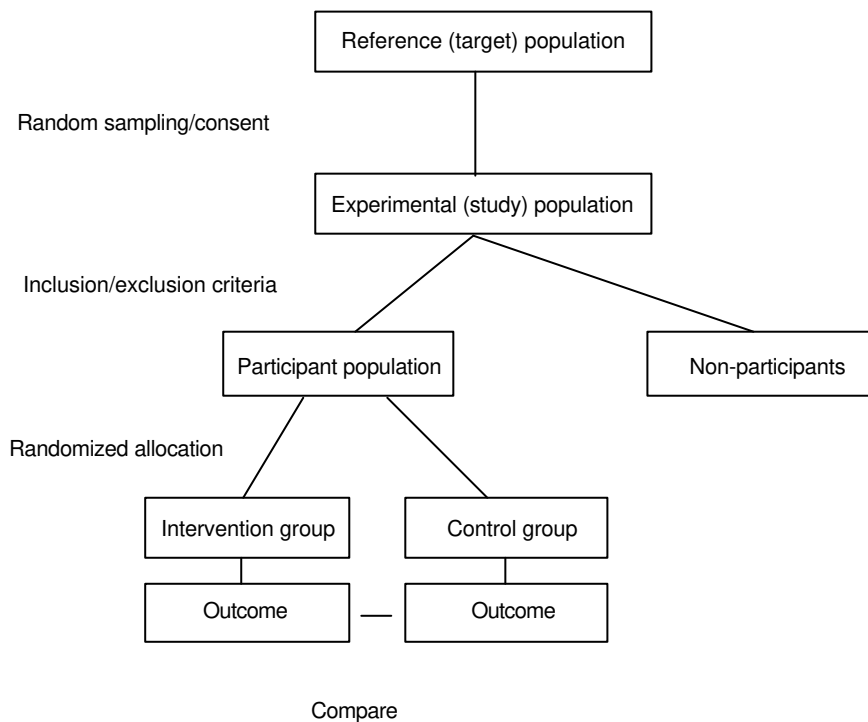
The reference population (also known as the target population) is the population to which generalizations of the results of the experiment apply. In a clinical trial of AZT for AIDS patients, all potential patients diagnosed with AIDS would be included in this group. In a community trial on the effect of maternal education in nutrition on improvements in children's health, all the families with children will form the reference population. The first step in any experimental design will be to identify the appropriate reference population.

Once the reference population is identified, one has to determine if this entire population is available for study, or if only a sample is available. If a sample, the best choice for this study population would be to obtain a simple random sample (especially if the study is reasonably large), so that there will be no bias in the selection of study subjects. In practice, however, the study group is chosen for convenience. For example, in the case of a clinical trial, the study is often restricted to a few sites and patients reporting to these sites, either voluntarily or through some referral system. Clearly, the success of the experiment, and whether the results of the experiment can be generalized to the target population, will depend on how representative this sample is. For example, if a clinical trial of cardiac patients were limited to a specialized cardiac care centre, the results would not be generalizable to all cardiac patients. Selection bias is often a very serious issue when choosing a study population.

In most experiments, the study may be restricted to a subgroup of the study population for various reasons. These are (and must be) listed prior to the beginning of the experiment as inclusion/exclusion criteria. The inclusion criteria identify the target group in a consistent and reliable manner. For example, if the experiments are done on patients suffering myocardial infarction, one must define myocardial infarction in a way that is acceptable to the receivers of the information from the study (usually other physicians or health care workers, policy-makers, etc.) and it must be identified as such in different settings. The definitions should be precise and reproducible. The exclusion criteria, on the other hand, list characteristics of the study subjects that would make them ineligible to enter the study. These are chosen to minimize potential dangers (elderly patients and pregnant women are generally excluded from clinical trials), and to select relatively homogeneous subjects to reduce the required sample size. Clearly, this will 'bias' the group subjected to experiment from the target or study group. The scope of exclusion due to these criteria may make the study invalid.

Once the potential group of subjects is determined, it is essential to get 'informed consent' from the participants before they are subjected to experiments. It is unethical to conduct an experiment on human beings without their consent (either consent of the individual or, if incapacitated or unable to give legal consent, consent by the authorized substitute). There are several international conventions (e.g. the Helsinki declaration) on human experimentation, and these should be consulted. In some cases, country-specific information and consent forms may be required to obtain proper consent. Subjects unwilling or unable to give consent should be removed from the list of potential participants.

The subjects left after the preceding exclusions form the study participants. This is the group that is to be the subject of experimentation. The subjects in this group will then be randomly allocated to the various intervention factors and the control group. The random allocation may be done using a simple random sample approach, or may be stratified according to various confounding factors.

Once the subjects are allocated to the experiment and control groups, they are followed for a specified period of time under strict conditions, and the outcome of the experiment is carefully documented. The outcome may be a dichotomous event such as a cure of the disease, relief of pain, etc., or it could be measured as a continuous variable, such as a reduction in blood pressure, or intervals of recurrence. The outcome measures are then compared between the groups using appropriate statistical methods.

It should be noted that, although the subjects have been randomly allocated to the experimental and control groups, there is no guarantee that the allocation will be free of bias. There could still be noticeable differences between the two groups with respect to one or more potential confounding variables. Therefore, success of the random allocation has to be verified by comparing the distribution of all confounding variables at the beginning of the experiment (prior to treatment allocation). If there is any appreciable difference, appropriate adjustments should be made when completing the statistical comparison of outcome measures.

4.4 The randomized clinical trial (RCT)

The most commonly encountered experiment in health science research, and the research strategy by which evidence of effectiveness is measured, is the randomized, controlled, double-blind clinical trial, commonly known as an RCT. This design follows the design illustrated in Figure 4.1. The RCT may be summarized in the following flow diagram.

FIGURE 4.2   DESIGN OF A RANDOMIZED, CONTROLLED, DOUBLE-BLIND CLINICAL TRIAL
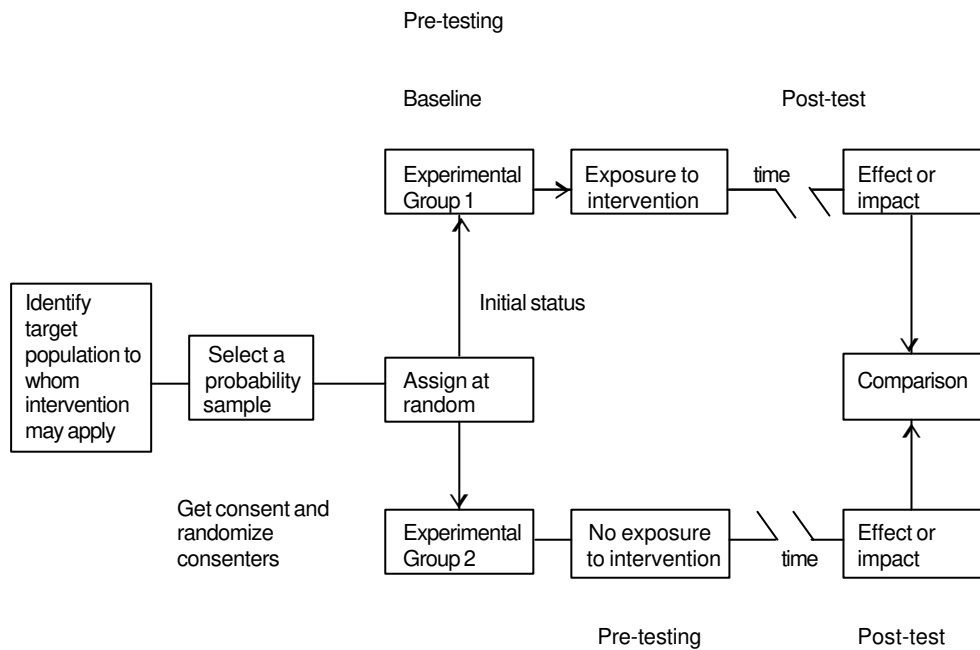
Clinical trials may be done for various purposes. Some of the common types of clinical trial (according to purpose) are:

a.   prophylactic trials, e.g. immunization, contraception;

b.   therapeutic trials, e.g. drug treatment, surgical procedure;

c.   safety trials, e.g. side-effects of oral contraceptives and injectables;

d.   risk-factor trials, e.g. proving the etiology of a disease by inducing it with the putative agent in animals, or withdrawing the agent (e.g. smoking) through cessation.

Therapeutic trials may be conducted to test efficacy (e.g. does a therapeutic agent work in an ideal, controlled situation?) or to test effectiveness (e.g. after having established efficacy, if the therapy is introduced to the population at large, will it be effective when having to deal with other co-interventions, confounding, contamination, etc.?)

The intervention in a clinical trial may include:

a.   drugs for prevention, treatment or palliation;

b.   clinical devices, such as intrauterine devices;

c.   surgical procedures, rehabilitation procedures;

d.   medical counselling;

e.   diet, exercise, change of other lifestyle habits;

f.   hospital services, e.g. integrated <u>versus</u> non-integrated, acute <u>versus</u> chronic care;

g.   risk factors;

h.   communication approaches, e.g. face-to-face communication <u>versus</u> pamphlets;

i.   different categories of health personnel, e.g. doctors <u>versus</u> nurses;

j.   treatment regimens, e.g. once-a-day dispensation <u>versus</u> three times a day.

Each of the scenarios will follow the same design as illustrated in Figure 4.1, but details of each step may be different. For example, randomization between the two experimental groups may pose different problems for the different types of interventions above. We will discuss a therapeutic trial in detail to examine the various steps and the issues that have to be dealt with.

Traditionally, clinical trials of new therapies or devices pass through four phases:

a. Phase I clinical trial

This first phase in humans is preceded by considerable research, including pharmacological and toxicological studies in experimental animals to establish that the new agent is effective and may be suitable for human use, and to estimate roughly the dose to be used in man. Phase I trials include studies of volunteers who receive, initially, a fraction of what the anticipated dose is likely to be, and are monitored for effects on body functions, such as hepatic, cardiovascular, renal, gastroinstestinal and endocrinal functions. The metabolism of the drug may also be investigated at this stage. These studies are normally done on volunteers, who are usually institutionalized, and occupy what are called 'research beds'. They require close supervision. This phase, which is of short duration (usually one or two months), requires high technology in biochemistry, pharmacology and endocrinology, and varied medical expertise. It also requires access to highly developed laboratory facilities.

b. Phase II clinical trial

This phase is also carried out on volunteers selected according to strict criteria. The purpose of Phase II is to assess the effectiveness of the drug or device, to determine the appropriate dosage, and to investigate its safety. Further information on the pharmacology, especially the dose-response relationship of the drug, is collected. In the case of a device, its effectiveness is assessed and its configuration is tested and, if needed, improved.

c. Phase III clinical trial

This is the classical phase (the one usually referred to as a 'clinical trial' and reported in health research journals). It is performed on patients, who should consent to being in a clinical trial. Strict criteria for inclusion in and exclusion from the trial are followed. The purpose of this phase is to assess the effectiveness (one could argue that it is still only an efficacy trial, because of the strict conditions under which the study is conducted) and to assess safety in continued use of the drug or device in a larger and more heterogeneous population than in Phase II. It includes more detailed studies and monitoring than those given in a normal service situation. This phase is usually carried out on hospital inpatients, but may be performed on outpatients

with intensive monitoring and follow-up. It requires superior clinical and epidemiological skills, in addition to the required laboratory technology. It also requires proper planning, organization and strict adherence to preformulated protocols and instructions, especially in multi-centre collaborative trials. Emphasis is also given to proper record keeping, follow-up and supervision.

Results from Phase III trials are used by regulatory agencies to evaluate whether a new product or device should be licensed for general public use. Initial Phase III trials therefore, have strict guidelines on the type and amount of data to be collected, the way the data are analysed and presented, and their dissemination to the users (patients and health care workers).

d. Phase IV trial

Although it has been customary to approve drugs or devices for general use following successful Phase III trials, increasing interest has been shown by governments, and by WHO and other agencies, in subjecting drugs and devices to yet another phase, i.e. a trial in normal field conditions. The purpose of the Phase IV trial is to re-assess the effectiveness, safety, acceptability and continued use of the drugs or devices under these conditions. Note that Phase III trials are often time-limited, and any adverse effects may not become apparent in such a short time. Phase IV trials add to the evidence of safety from this perspective. They also encompass a formulation of the service requirements of the new method, including facilities, training, logistics of supply and transportation, supervision, and other programme aspects. Although this phase is carried out under conditions that are as close to normal as possible, Phase IV requires additional epidemiological and biostatistical skills, as well as research requirements, including record-keeping and computer facilities.

4.5 Factors that influence the design and analysis of clinical trials (Phase III)

a. The agent, treatment or experimental factor:

Knowledge that is as complete as possible about the treatment should be available to the researchers. This knowledge usually comes from Phase I and Phase II trials, as well as from many ancillary sources. For example, one should know the pharmacological action, toxicity, dose, safety and method of administration of the drug.

b. *Conditions to be treated:*

Adequate clinical and epidemiological knowledge about the conditions to be treated should be available to researchers. This includes the natural history of the condition, the diagnostic criteria, the routine medical management, and other variables that can influence the progress of the condition (age, sex, social conditions, weight, smoking status, co-existing conditions, drugs taken for other reasons). Detailed treatment procedures should be explicitly stated and adhered to.

c. *The target population:*

The type of cases to be included in the trial should be carefully specified, with explicit criteria for inclusion and exclusion from the trial. The size of samples should be predetermined. If one institution cannot provide the required sample, collaborative trials should be carefully planned, with rigid protocols. A single-centre trial has the advantage of subject homogeneity, while a multi-centre trial would be more generalizable. It should also be noted that in multi-centre trials, ensuring the comparability of the quality and quantity of data collected between centres is much more difficult than if the trial is limited to one centre where the principal investigator is based. Informed consent should be obtained before cases are assigned to experimental and control groups. Strict procedures should be used in allocating cases to groups.

d. *Ethical issues:*

No clinical trial should ever be performed without due consideration of ethical issues. Usually the protocol is subjected to ethical review in-house, or by an independent review board, and only after such approval should the trial begin. Most regulatory agencies will insist on the ethical approval of the trial before they will consider the results from the trial as evidence of efficacy or effectiveness.

e. *Outcomes to be measured:*

One should specify explicitly what outcomes are expected, and what criteria are to be applied to determine the success or failure of the trial. The outcomes may include prevention of a condition, cure of a condition, improvement in the condition, alleviation of pain, improved physical or mental health.

f. Side-effects:

Criteria for observing and recording side-effects should also be formulated. If side-effects would endanger the health of a patient, he/she should be excluded from the study and treated appropriately. Procedures for stopping the trial if too many unwanted side-effects are observed, should be incorporated into the study design.

g. Study instruments:

These are also to be specified, to include laboratory tests, clinical diagnosis, procedures, special interview inventories and questionnaires, or use of proxy information (from spouses, relatives, neighbours, treating physicians) for obtaining medical and social histories.

h. Blinding:

It is highly desirable to enhance the objectivity of measurements by 'blinding' or hiding the identity of the treatment from the subject, from the investigator who evaluates the outcome, and sometimes from the person who enters and analyses the data. When the investigator and the patient are both blinded, it is known as a 'double-blind' study, and is the most common form of clinical trial. The treatment allocation, however, should be known to a select committee independent of the investigators, who will monitor the progress of the study and stop the trial if (a) the study arrives at a justifiable conclusion long before the trial sample size is reached, or (b) an unacceptable level of side-effects has been uncovered.

In some cases, blinding of all parties may not be feasible. For example, in most surgical trials, or trials involving medical devices, it may be obvious to which group the patient has been allocated. Even in the case of surgery, sham surgical procedures (akin to a placebo for drug trials) have been employed in some studies.

i. Stopping rule:

Criteria for terminating the trial should be clearly specified. In most cases, a fixed sample size has to be reached before the trial is stopped. A variation on this is the sequential clinical trial, where the results are analysed frequently, and the trial

stopped as soon as statistically significant differences are observed. (Clearly, there will be a penalty for this 'frequent look at the data', in the form of adjusted type I errors.) Procedures for terminating the trial prematurely because of adverse effects should also be specified and adhered to.

j. *Plans for analysis:*

No clinical trial should be undertaken in the absence of epidemiological and statistical talent on the research team. Detailed plans for analysis must be made prior to the trial. It is unethical to subject people to experimentation without having adequate competence in the research design and analysis of data.

k. *Selective attrition:*

A very serious threat to the clinical trial is the attrition of patients after they have been randomized into the trial (before actually beginning the treatment, or after having undergone partial treatment). This will reduce the power of the study (reduced sample size) and increase the chances of bias (those who drop out may be different from those in the trial). Therefore, it is imperative to ensure that the attrition is small.

Because some of the reasons for attrition (e.g. severe deterioration of condition, late discovery of misdiagnosis or non-applicability of diagnostic criteria) may be justifiable, it is usually the practice to inflate the initial estimate of sample size by 10%, so that its power is not drastically reduced. The question of bias can only be answered after the fact, by comparing the drop-outs with those completing the trial on important co-variables. In any case, if the attrition rate is large, the results of the trial will be suspect.

l. *Methods for ensuring the integrity of the data:*

It goes without saying that impeccable record keeping is absolutely essential in research, but this is particularly so in the case of clinical trials. Apart from the integrity of the research results, there are often regulatory requirements and legal obligations to keep all the data, and to guarantee adequate supervision of data collection, quality control, analysis and reporting.

m. The choice of design:

There are various experimental designs for clinical trials. The choice depends on the nature of the trial components and the composition of the research team. The usual design is the randomized, controlled, double-blind clinical trial (see Figure 4.4). Forms other than those described here are available, e.g. (i) multiple-dose design, (ii) multiple-treatment design, (iii) sequential design, (iv) factorial design, and (v) various forms of 'blocked' designs, such as the Latin square, balanced incomplete block design, cross-over trials, etc.

n. Time required:

One should allow several months for planning the trial, to include: preparation of protocols, sampling procedures, determination of sample size, identification of sources of subjects, outlining management procedures including quality control, planning and analysis of data. Sometimes a feasibility study may be necessary before the trial to test out the protocol and determine what is possible and what is not.

4.6 Community intervention trials (CITs)

CITs are usually carried out in hospitals or clinics, and are usually directed at a patient group with specific health conditions. However, randomized experiments are also sometimes done in the community. The classic example of a community intervention trial would be that of testing a vaccine. Some communities will be randomly assigned to receive the vaccine, while other communities will either not be vaccinated, or will be vaccinated with a placebo. Another example would be a test of whether the introduction of iron-fortified salt in the community would reduce the incidence of anaemia in the community.

In these types of studies, the major difference from the RCT is that the randomization is done on communities rather than individuals. Communities selected for entry to the study have to be similar as much as is possible, especially since only a small number of communities will be entered.

Very often, blinding is not possible in these types of studies, and contamination and co-interventions become serious problems. Contamination occurs when individuals from one of the experimental
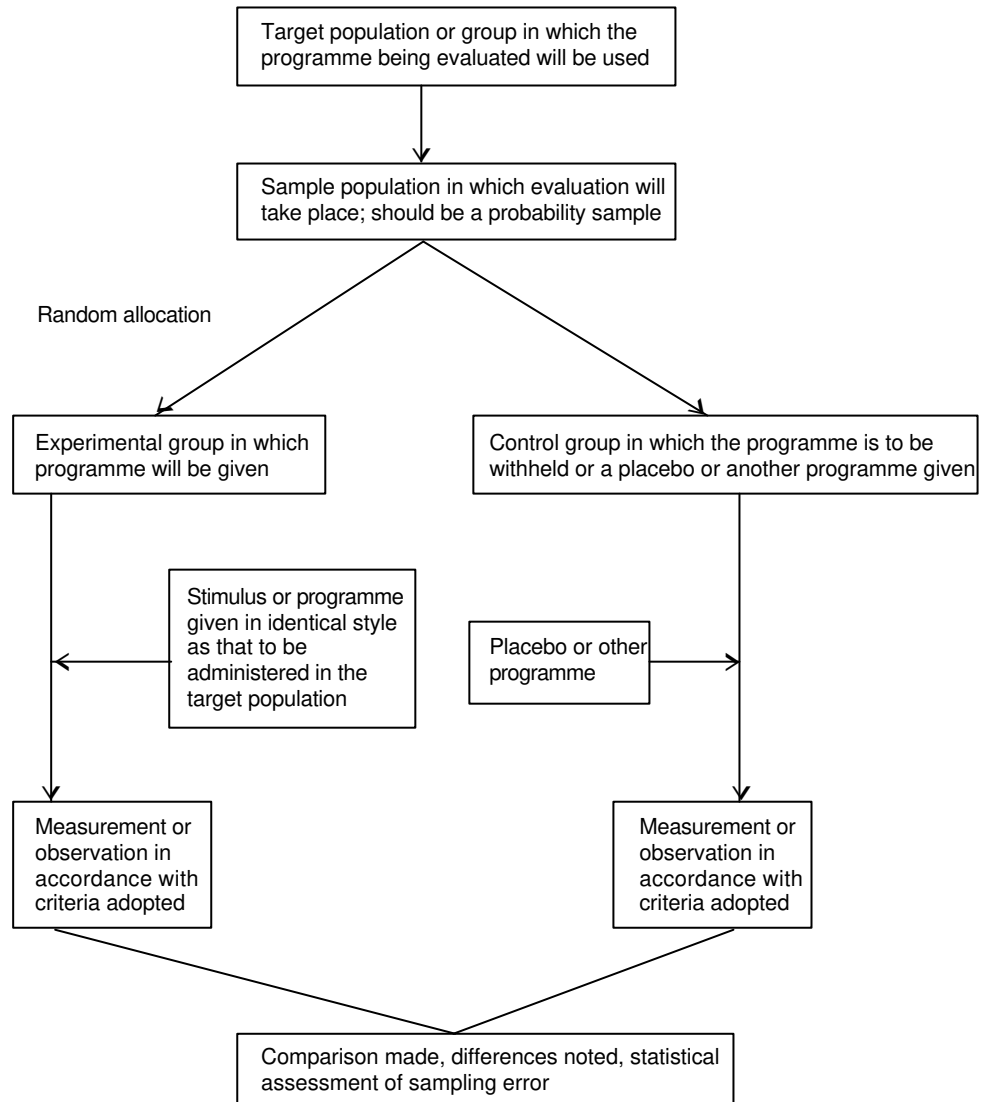
groups receive the intervention from the other experimental group. For example, in the study of iron-fortified salt, some of the members of the community receiving non-fortified salt might hear about the fortified salt, and may acquire it from the other community. (The reverse is also possible.) This is particularly so if the communities are geographically close. Co-intervention occurs when other interventions, either unknown to the investigators of this trial or otherwise, are simultaneously introduced, in which case, comparison of results from the two randomized groups will no longer be a reflection of the intervention under trial. The fact that these trials use randomization by communities also reduces the sample size; the effective sample size is the number of communities, not the number of people in these communities. Special statistical procedures have to be applied to take into account this 'clustering effect'.

Most of the community intervention trials involve evaluative strategies to study community health services. Typical examples of such trials involve:

·   evaluating the need for a service, i.e. community diagnosis (assessment or evaluation of needs);

·   evaluating the design of a health service (design evaluation);

·   evaluating the performance or efficiency of the process of delivery of the services (efficiency or process evaluation);

·   evaluating the effectiveness and impact of the programme or procedure (effectiveness or impact evaluation);

·   relating the outcome to the input and constraints of the programme (system evaluation) including cost-benefit analysis.

Figure 4.3 shows the necessary steps in the organization of a community trial.

FIGURE 4.3 PROCEDURES IN A FIELD TRIAL

```
                    ┌─────────────────────────────────┐
                    │ Target population or group in    │
                    │ which the programme being        │
                    │ evaluated will be used           │
                    └─────────────────────────────────┘
                                    │
                                    ▼
                    ┌─────────────────────────────────┐
                    │ Sample population in which        │
                    │ evaluation will take place;       │
                    │ should be a probability sample    │
                    └─────────────────────────────────┘
                           ╱              ╲
   Random allocation      ╱                ╲
                         ╱                  ╲
        ┌──────────────────────┐   ┌────────────────────────────────┐
        │ Experimental group in │   │ Control group in which the      │
        │ which programme will  │   │ programme is to be withheld or  │
        │ be given              │   │ a placebo or another programme  │
        └──────────────────────┘   │ given                           │
               │                    └────────────────────────────────┘
               │                                  │
    ┌──────────────────────┐       ┌──────────────────┐
    │ Stimulus or programme │       │ Placebo or other  │
    │ given in identical    │ ◄──   │ programme         │  ──►
    │ style as that to be   │       └──────────────────┘
    │ administered in the   │
    │ target population     │
    └──────────────────────┘
               │                                  │
               ▼                                  ▼
    ┌──────────────────┐           ┌──────────────────┐
    │ Measurement or    │           │ Measurement or    │
    │ observation in    │           │ observation in    │
    │ accordance with   │           │ accordance with   │
    │ criteria adopted  │           │ criteria adopted  │
    └──────────────────┘           └──────────────────┘
                  ╲                 ╱
                   ╲               ╱
            ┌─────────────────────────────────────┐
            │ Comparison made, differences noted,  │
            │ statistical assessment of sampling   │
            │ error                                │
            └─────────────────────────────────────┘
```

Example

A study was designed to evaluate the impact of increased screening for cervical cancer on mortality from that disease. Six medium-sized, comparable cities in the southern region of the USA were assigned at random to three clusters of two cities each, as follows:

Cluster A: receives an intensive neighbourhood campaign of education, plus messages through the mass media to motivate women aged >30 to join the screening programme for cancer of the cervix.

Cluster B: receives messages through the mass media only.

Cluster C: receives no special programme beyond routine services. This cluster is used as a blank control, and satisfies the need to know what would happen with regard to mortality from cancer of the cervix if no special education programmes were instituted to increase utilization of the screening services.

Notice that mass media were used as one of the methods for dissemination of information. This has the potential to introduce contamination into the control community, especially if the visual medium of TV is used, as the programmes may be broadcast in the control community.

The criterion of success in this evaluation study (which is experimental) is the relative reduction in the annual rate of mortality from cervical cancer in the three clusters. This necessitates measuring the mortality before (average for the three years preceding the intervention) and after the study. In order to mitigate ecological fallacy in comparing the three clusters (since reduction could have been due to reasons other than the screening service), a record is kept of the number and characteristics of the women using that service. In addition, private practitioners agree to report any screening service they provide. The number of users is utilized only as an indication that the change in mortality could have been due to screening.

The study was continued for two years. During this time, the relative reduction in mortality rates for cervical cancer was highest in cluster A, and intermediate in cluster B. There was a parallel increase in the number of users of the screening services, with the greatest increase in cluster A, followed by cluster B, and the smallest in cluster C.

4.7 References and further reading

Altman D.G. Statistics and ethics in medical research. British Medical Journal, 1980, 281: 1336-1338.

Angell M. The ethics of clinical research in the third world. New England Journal of Medicine, 1997, 337(12): 847.

Berry A.D. Statistical issues in drug research development. New York, Marcel Dekker, 1990.

Buncher R.C., Tsay, J.Y. Statistics in the pharmaceutical industry. New York, Marcel Dekker, 1981.

Friedman L.M., Fuberg C.D., DeMets D.L. Fundamentals of clinical trials. Boston, Wright, 1983.

Gore S.M. Assessing clinical trials: why randomize? British Medical Journal, 1981, 282: 1958-1960.

Lilford R.F., Jackson J. Equipoise and the ethics of randomization. Journal of the Royal Society of Medicine, 1995, 88(10): 552-559.

Peto R. et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. British Journal of Cancer, 1976, 34: 585-612.

Peto R. et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. British Journal of Cancer, 1977, 35: 1-39.

Rothman K.J., Michels K.B. The continuing unethical use of placebo controls. New England Journal of Medicine, 1994, 331(6): 394-398.

Chapter 5

Sampling Methods and Sample Size

5.1 Introduction

Most research studies involve the observation of a sample from some predefined population of interest. In epidemiological studies, for example, a sample of people is observed for exposure to various risk factors, health outcomes and other related variables. The conclusions drawn from the study are often based on generalizing the results observed in the sample to the entire population from which the sample was drawn. Therefore, the accuracy of the conclusions will depend on how well the samples have been collected, and especially on how representative the sample is of the population. In this chapter, we will discuss the major issues that a researcher has to face in selecting an appropriate sample.

5.2 Why sampling?

Sampling is a process of choosing a section of the population for observation and study. There are several reasons why samples are chosen for study, rather than the entire population. First and foremost, a researcher wants to minimize the costs (financial and otherwise) of collecting the information, processing this information and reporting on the results. If a reasonable picture of a population can be obtained by observing only a section of it, the researcher economizes by choosing such a section of the population. Obviously, when a sample is observed, the total information will be less than if one were to observe the entire population.

However, in some cases, the process of observing the entire population would take such a large amount of time and resources that (a) the results would not be timely, and (b) the observations might be less reliable. Consider the common approach to observing the entire population, the census. Most countries collect information on their population periodically (every five years, every ten years, etc.) through census. This involves enumerating every individual in the population and collecting a predetermined set of information. Even in a relatively small country such as Canada (population, 29 million), the process takes a substantial part of a year, and the tabulated observations are not available for several years after the census. When the population size is large, for example in India or China, the data analysis and reporting may be delayed even further. In addition, the census is never able to collect information on all the population: the homeless and nomadic sections of the population are often missed.

A major advantage of sampling over complete enumeration is the fact that the available resources can be better spent in refining the measuring instruments and methods so that the information collected is accurate (valid and reliable). Some information, such as monitoring of the body burden of toxic metals in the population, which may require specialized equipment and staff, cannot be collected from the entire population. A sample in such cases would provide a reasonable picture of the population status.

5.3   Process of sampling

What determines a proper sample? The primary concern in selecting an appropriate sample is that the sample should be representative of the population. Every variable of interest should have the same distribution in the sample as in the population from which the sample is chosen. This requires knowledge of the variables and their distribution in the population, which of course is why we are doing the study in the first place! Therefore, it is not often possible to ensure the representativeness of the population. However, statisticians have come up with ways in which we can give a reasonable guarantee of representativeness. We will discuss some of these methods briefly in later sections.

Before a sample is drawn, the population has to be clearly defined. In a population survey, this requires having a list (sampling frame) of all the individuals in the population. Probabilistic methods can then be developed to draw a sample in such a way that we can

assure the representativeness of the various characteristics in which we are interested. In experiments (such as clinical trials) this list may not be explicit, and may evolve as the sampling progresses. For example, a list of inclusion and exclusion criteria would be specified at the beginning of the trial, defining the general framework for the population. Then, as patients are identified, they will be selected for study, and allocated to various experimental groups using probabilistic methods.

The sampling frame consists of a list of elements (units) of the population. In population surveys, this is a list of people. In clinical trials for a disease, it is a list of patients with that disease. In a case-control study, it is a list of people with the disease and a list of people without the disease. The completeness and accuracy of this list is essential for the study to be successful. One of the major flaws in many research projects is a biased selection of the sampling frame. For example, if a telephone survey is conducted in India before a general election to predict which party will win, the results will most likely be wrong, since the sampling frame consists of only affluent people (who can afford a telephone), and their opinions are not likely to be representative of the entire population.

Once a sampling frame has been identified, one needs to have methods of selecting individuals from this frame to be included in the study. Two issues are important: how large a sample should be selected, and how the individual units should be selected. These issues are discussed in the following sections.

5.4 How large a sample?

One of the most difficult decisions facing the researcher is how large his sample should be. Two common approaches are employed in research studies: the empirical and the analytical. The empirical approach involves using sample sizes that have been used in similar studies. This has no scientific basis, and will only be satisfactory if the previous studies had acceptable limits on the errors of generalization, and the current study is very similar in its scope (objectives, design, study population, etc.). This method is not recommended and will not be discussed further.

The analytical (scientific) approach to determining the appropriate size of the sample to be included in the study depends on the assessment of errors of inference, and a desire to minimize 'sampling error'. Sampling error measures the amount of variability between

sample results (as a proxy for closeness to the real situation in the population, and as reproduced in the sample results); the less variable the sample results are, the closer the sample results are to the population results.

The main determinant of the sample size is, therefore, how accurate the results need to be. This depends on the purpose of the study (descriptive study to determine a summary measure of a characteristic, or an analytical study where specific sets of hypotheses are being tested).

5.4.1 Sample sizes for descriptive studies

In the case of descriptive studies, often the object is to obtain an estimate of a population parameter. For example, in opinion polls, the market researcher may be interested in finding out what proportion of people prefer a particular brand. A nutritionist may be interested in the average daily caloric intake of the population. A health researcher may be interested in the proportion of people who smoke, or the median survival after coronary bypass surgery. The determination of the size of sample required to answer these questions depends on several factors:

i.   What is the measure of interest?   This would have been determined by the study objectives.   The identification of the characteristic of primary importance determines the next steps in the process of defining the sample size. For example, if a prevalence rate in the population is to be estimated by observing a sample from the population, the measure is the proportion of people in the sample with the disease.

ii.   What is the underlying probability distribution of the characteristic of interest? Most research questions fall into one of two possible scenarios: the binomial distribution (when one wants to estimate the proportion of a certain event), and the normal distribution (when one wants to estimate an average value).   The market researcher above, for example, has the preference of a brand as the characteristic, with two possible outcomes.   If one assumes that there is possibly a fixed proportion (π) of people with preference for the brand, then the number of people expressing this preference in any fixed set of people will follow a binomial distribution, with the proportion (p) of the people showing the preference as a good estimate of the population proportion. For the nutritionist, the daily caloric intake of individuals follows a normal distribution with some

average (μ), and the average of the daily caloric intake of the sample of people (x) observed would be a good estimate of this population value.

iii. What is the sampling distribution of the measure? Drawing inferences from the sample to the population involves inherent errors, which are measured by the sampling distribution. If we observed several samples, under the same method of selecting the samples, the measures from each of these samples would vary, resulting in a 'probability distribution' for the sample measure. This distribution is called the sampling distribution, and it depends on the type of study design and on how the samples were obtained. In calculating sample sizes, it is often assumed that the sampling involves simple random sampling (discussed later in this chapter). Sometimes the sampling design is much more complicated (e.g. multistage cluster sampling techniques) and more complicated formulae will have to be used to calculate sample sizes appropriately.

iv. How accurate do you want the results to be? Basically, one is interested in obtaining an estimate as close to the population value as possible. Therefore, some measure of the difference between the estimate and the population value has to be considered. In most cases, a mean-squared error (average of the squared deviation of the sample value from the population value) is used. A concise way of expressing this error is to use the 'standard error of the estimate'. The standard error comes from the sampling distribution of the estimate. If the sampling is done properly (with appropriate probabilistic methods), one can predict what this distribution should be, and based on this, one can estimate how close to the population value the sample estimate will be:

For example, in the case of estimating the population proportion, the sampling distribution of the sample proportion, p is approximately normal, with mean π and variance $\pi(1-\pi)/n$, where n is the sample size. This gives the ($1-\alpha$) confidence interval for π to be

$$p \pm z_{1-\alpha} \sqrt{p(1-p)/n}$$

where $z_{1-\alpha}$ is the appropriate cut-off point on the standard normal distribution. (For example, for 95% confidence, $z_{1-\alpha} = 1.96$.)

The accuracy of the estimate therefore depends on two quantities: how narrow this interval is (width of the interval) and how confident we are (e.g. 95%).

The calculation of the size of the sample for a descriptive study therefore depends on the two parameters – the width of the confidence interval and the confidence coefficient. Computer programs are readily available (e.g. EPIINFO has a module that allows for the computation of sample sizes). The two common scenarios, estimating a population proportion and estimating a population mean, are illustrated below:

a.   Estimating a population proportion (p). Suppose we want to conduct a survey to determine the prevalence (π) of a relatively common disease in a community. We want to determine how many people should be observed to obtain a reasonably accurate picture of the prevalence. The following steps are necessary:

·   Specify the parameters of error:

Confidence coefficient (1-α)   95%

Width of the interval (δ)   10%

·   Make a guess as to the value of π 30%

The problem is to calculate the sample size required for estimating the prevalence of the disease within $\pm 5\%$ of the true value, with 95% confidence. Since the confidence interval actually depends on the true value, p, we have to make a guess as to what this value might be. This is done based on prior experience; if no guess is available, use the value 50%, which will give the largest sample size. Using the fact that the sample proportion (p) has the confidence interval given above, the sample size (n) can be calculated using the formula:

$$n = (z_{1-\alpha} / \delta)^2 \, p(1-p)$$

In the above example, therefore, n = $(1.96/5)^2(30*70)$ = 323; we need a minimum of 323 subjects observed to assure that the 95% confidence interval for the estimated proportion will be within 5% of the true prevalence. If the true prevalence is less than 30%, the confidence interval will be narrower. The maximum sample size required will occur when the true prevalence is 50%, in which case, n = 385.

The above calculation assumes a simple random sample from a relatively large population. In practice, the population from which the samples are drawn may be fixed and small, in which case corrections to the above formulae are required. (See EPIINFO program for variations of this formula, and use under different sampling designs.)

b. Estimating a population average (μ). Suppose we want to estimate the average daily caloric intake of people in a community. The daily caloric intake is assumed to have a normal distribution around μ, with a standard deviation (σ). The sample measure used to estimate μ is the sample mean. The sampling distribution of the sample mean is also normal, with the same mean, μ and standard deviation, $\sigma/\sqrt{n}$ (the standard error of the mean). Notice that we need to know the value of σ to proceed further. It is either obtained from other similar studies, or by actually obtaining a small number of observations at random in a test study. If neither of these is possible, one may make a reasonable guess by taking the maximum range (maximum value possible – minimum value possible) and dividing this range by 4. (Using the supposition that for normal distribution, 95% of values will be within ± 2 standard deviation from the mean, and the mean will be the central value.) Then the following steps will help calculate the sample size:

· Specify error parameters:

Confidence coefficient (1-α): 95%

Width of the interval (δ): 50 cal.

· Obtain the standard deviation (σ): 150 cal.

· The 95% confidence interval for the sample mean is:

$$\bar{x} \pm z_{(1-\alpha)}\sigma / \sqrt{n}$$

· Therefore the required sample size in the example is:

$$n = (1.96*150/50)^2 = 35.$$

c. Estimating relative risks or odds ratios. The formulae for calculating sample sizes in these situations are much more complicated, since the sampling distribution of the estimates of relative risks and odds ratios are not simple. Various computer programs are available to calculate the appropriate sample sizes.

The principles are essentially the same: determine the formula for confidence interval, and by specifying the two parameters, calculate the sample size from this formula.

5.4.2 Sample sizes for analytical studies

Since the primary purpose of an analytical study is to test (one or more) null hypotheses, the determination of the sample sizes requires the specification of the limits of errors one is willing to accept in accepting or rejecting the null hypothesis (type I and type II errors). As in the case of descriptive studies, one has to determine the sample measures used (a proportion, a sample mean, an estimate of RR or OR, etc.) and their sampling distribution (on the basis of which, a decision to accept or reject null hypothesis is taken). By equating the two types of errors based on the sampling distribution to the pre-set limits on these errors, we can work out the sample size.

For example, suppose we decide to accept a type I error, or α (probability of making a false conclusion that the two proportions are not equal in the population, when they are in fact equal). The calculation of a type II error, or β (probability of making a false decision that the two proportions are equal when they are not) depends on a precise definition of 'null hypothesis is not true'. The simplest way to do this is to define the smallest difference (δ) in the two proportions that we consider meaningful (clinically significant difference) and calculate β under this hypothesis. Clearly, if the difference is larger than δ, the probability of type II error will be less. Using this approach, formulae have been derived for calculating sample sizes for various types of statistical tests. [Note: In statistical tests, the discussion of type II errors may be worded in terms of 'statistical power', which is simply $1-\beta$: i.e. having a 5% type II error is the same as the study having 95% 'power'.] The more common of these situations are summarized below. (As before, computer programs are readily available for most of these cases, and the computation here is presented solely for illustrative purposes.)

a. Testing equality of two proportions: $\pi_1 = \pi_2$.

The sample measures used are the sample proportions, and the sampling distribution used in testing this null hypothesis is either the standard normal distribution (z), or equivalently the chi-square (χ^2).

· Set type I error: α;

· Determine 'minimum clinically significant difference': δ;

· Make a guess as to the 'proportion' in one group (usually 'control'): π_1 ;

· Determine the power required to detect this difference: $(1-\beta)$.

The sample size required is:

$$n = [\{ z_{1-\alpha} \sqrt{2\overline{\pi}(1-\overline{\pi})} - z_{\beta} \sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)} \}/\delta]^2$$

where $\overline{\pi} = (\pi_1 + \pi_2)/2$

For example, suppose we are interested in determining the sample size required in a clinical trial of a new drug that is expected to improve survival. Suppose the traditional survival rate is 40%, i.e. $\pi_1 = 0.4$. We are interested in detecting whether the new drug improves survival by at least 10%, i.e. $\delta = 0.10$, therefore $\pi_2 = 0.50$. Suppose we want a type I error of 5%, i.e. $\alpha = 0.05$, therefore $z_{1-\alpha} = 1.96$; we also want the type II error (β) to be 5%, or we want to detect a difference of 10% or more with a probability of 95%: therefore $z_{\beta} = -1.645$.

Substituting these values in the above equation gives n = 640. Thus the study would require 640 subjects in each of the two groups to assure a probability of detecting an increase in the survival rate of 10% or more with 95% certainty, if the statistical test used 5% as the level of significance.

b. *Sample size for a case-control study*

Suppose that long-term use of oral contraceptives (OC) increased the risk for coronary heart disease (CHD) and that one wished to detect an increase in relative risk of at least 30% (equivalently, OR>1.3) by means of a case-control study. What would be the proper sample size?

The test of hypothesis in the study will be equivalent to testing if the proportion of women using OC is the same among those with CHD and those without CHD. We need to determine what proportion of women without CHD (controls) use OC; let us say 20%. Then we decide what will be the minimum difference that should be detected by the statistical test. Since we need to detect an OR>1.3, this translates to an increased use (24.5%) among the CHD patients, to give a difference of 4.5% to be detected. Choosing α and β to be 5% each, the sample size, using the above formula, would be 2220, i.e. we need to study 2220 cases and 2220 controls for the disease.

Sometimes the ratio of cases and controls may not be one-one, e.g. when the disease is rare, the number of cases available for study may be limited, and we may have to increase the number of controls (1-2, 1-3 etc.) to compensate. In such cases, the calculation of the sample size will incorporate these differences. Computer programs such as EPIINFO allow for these variations.

c. *Comparison of two population means*

When the study involves comparing the means of two samples, the sample measure that is used is the difference of the sample means. This has an approximately normal distribution. The standard error of difference depends on the standard deviations of the measurements in each of the population, and depending on whether these are the same or different, different formulae have to be used. In the simplest (and most commonly used) scenario, the two standard deviations are considered to be the same. We will illustrate the procedure.

We need to determine, as in case a, the minimum difference (δ) in the means that we are interested in detecting by statistical test: the two types of statistical errors (α and β) and the standard deviation (σ). Then the sample size required is calculated using the following formula:

$$n = [(z_{1-\alpha} - z_{\beta})\sigma / \delta]^2$$

For example, suppose we want to test a drug that reduces blood pressure. We want to say the drug is effective if the reduction in blood pressure is 5 mm Hg or more, compared with the 'placebo'. Suppose we know that systolic blood pressure in a population is distributed normally, with a standard deviation of 8 mm Hg. If we choose $\alpha = 0.05$ and $\beta = 0.05$, the sample size required in this study will be: $n = [(1.96+1.645)8/3]^2 = 34$ subjects in each group.

If the design is such that the two groups are not independent (e.g. matched studies or paired experiments) or if the standard deviations are different for the two groups, the formulae should be adjusted accordingly.

d. Comparison of more than two groups and multivariate methods

When considering sample size calculations for studies involving comparison of more than two groups, either comparing proportions or means, several other issues (e.g. which comparison is more important than the others: whether errors of paired comparison, or for the study as a whole are more important, etc.) have to be taken into account. Accordingly, the formulae for each of these situations will be much more complicated.

In multivariate analyses, such as those using multiple linear regression, logistic regression, or comparison of survival curves, simple formulae for the calculation of sample sizes are not available. Some attempts at estimating sample sizes using nomograms, or by simulating experiments and calculating sample sizes based on these simulated experiments, have recently appeared in the statistical literature. We will not discuss these here. When planning experiments, one of the crucial steps is in deciding how large the study should be, and appropriate guidance should be sought from experts.

5.5   Sampling methods

Once the population has been identified and the size of the sample determined, we need to decide how we are going to choose the sample from the population. [The size of the sample will also depend on this choice and therefore, the issue of sample size may have to be revisited after the choice of the sampling method; most of the discussions in the earlier section on sample size assumed a simple random sample.]

a. Simple random sample

This is the most common and the simplest of the sampling methods. In this method, the subjects are chosen from the population with equal probability of selection. One may use a random number table, or use techniques such as putting the names of the people into a hat and selecting the appropriate number of names blindly. Recently, computer programs have been developed to draw simple random samples from a given population. The simple random sample has the advantages that it is easy to administer, is representative of the population in the long run, and the analysis of data using such a sampling

scheme is straightforward. The disadvantage is that the selected sample may not be truly representative of the population, especially if the sample size is small.

b. Stratified sampling

When the size of the sample is small and we have some information about the distribution of a particular variable (e.g. gender: 50% male/50% female), it may be advantageous to select simple random samples from within each of the subgroups defined by that variable. By choosing half the sample from males and half from females, we assure that the sample is representative of the population with respect to gender. When confounding is an important issue (such as in case-control studies), stratified sampling will reduce potential confounding by selecting homogeneous subgroups.

c. Cluster sampling

In many administrative surveys, studies are done on large populations which may be geographically quite dispersed. To obtain the required number of subjects for the study by a simple random sample method will require large costs and will be inconvenient. In such cases, clusters may be identified (e.g. households) and random samples of clusters will be included in the study; then every member of the cluster will also be part of the study. This introduces two types of variations in the data – between clusters and within clusters – and this will have to be taken into account when analysing data.

d. Multi-stage sampling

Many studies, especially large nationwide surveys, will incorporate different sampling methods for different groups, and may be done in several stages. In experiments, or common epidemiological studies such as case-control or cohort studies, this is not a common practice. For details of these methods, see Levy and Lemeshow.

5.6   References and further reading

Cochran W.G. Sampling techniques. New York, John Wiley and Sons, 1977.

Fleiss H. Statistical methods for rates and proportions. New York, John Wiley and Sons, 1981.

Kish L. Survey sampling. New York, John Wiley and Sons, 1965.

Levy P.S., Lemeshow S. Sampling of populations: methods and applications. New York, John Wiley and Sons, 1991.

Sample size determination: a user's manual. Geneva, World Health Organization, 1986 (WHO/HST/ESM/86.1).

Schlesselman J.J. Case-control studies. Oxford, Oxford University Press, 1982.

Yates F. Sampling methods for censuses and surveys. London, Charles Griffin, 1981.

Chapter 6

Bias and Confounding

6.1   Introduction

As mentioned in Chapter 2, it is important to consider two sources of error when planning research: random error and bias. Bias occurs when the results of a study are systematically different from 'truth'. For example, if the objective of the study is to estimate the risk of disease associated with an exposure, and the result from the study consistently overestimates the risk, the result is said to be biased. Bias should be distinguished from random error, in that random error cannot be associated with a particular cause and tends to 'average out' in repeated sampling. Bias, on the other hand, would repeat the same direction of error in repeated sampling with the same design. Bias results from faulty design. There may be many reasons for bias, and care has to be taken to minimize bias when designing the study, since it is often difficult to separate the true effects from bias. Simply increasing the sample size, on the other hand, can minimize the effect of random error.

6.2   Types of bias

Several types of bias exist in research.   Sackett et al. have listed 19 types of bias commonly encountered in epidemiological studies. Choi has expanded this list further to 65. Indeed, any type of error introduced into the study, for which a cause can be identified, could potentially be considered a bias by definition (systematic error). Many of these are hard to detect and even harder to avoid. We shall consider three specific types of biases, which are very common in health research.

6.2.1 Selection bias

Selection bias is a distortion of the estimate of effect resulting from the manner in which the study population is selected. This is probably the most common type of bias in health research, and occurs in observational, as well as analytical studies (including experiments).

a. Prevalence-incidence bias

This type of bias can be introduced into a case-control study as a result of selective survival among the prevalent cases. In selecting cases, we are having a late look at the disease; if the exposure occurred years before, mild cases that improved, or severe cases that died would have been missed and not counted among the cases. This bias is not often a problem in cohort studies and experiments, but is quite common in case-control studies.

Example:

The high case-fatality rate in the early stages of clinically manifested coronary artery disease may invalidate the study of possible etiological factors, since the persons available for study as cases are the survivors (severe cases are absent). Likewise, myocardial infarction may be silent. Clinical features may be absent, and the biochemical and electrocardiographic changes in myocardial infarction may return to normal after an infarct (these mild cases will not appear among cases for study). The type of bias introduced into the study may be clear by contrasting a cohort study (where the disease is identified in all its forms) as shown in Table 6.1.

TABLE 6.1 COHORT VERSUS CASE-CONTROL STUDY: ESTIMATES OF THE RELATIVE ODDS OF DEVELOPING CORONARY HEART DISEASE AMONG MEN WITH AND WITHOUT CHOLESTEROLAEMIA

	Cohort Study			Case-control study		
	Developed CHD	Did not develop CHD	Total	CHD present	CHD absent	Total
Highest quartile of serum cholesterol	85	462	547	38	34	72
Lower three quartiles of serum cholesterol	116	1511	1627	113	117	230
Total	201	1973	2174	151	151	302
Odds ratio	2.40			1.16		

b. *Admission rate (Berkson's) bias*

This type of bias is due to selective factors of admission to hospitals, and occurs in hospital-based studies. Many case-control studies collect cases from hospitals, and identify controls from among patients in the hospital admitted for unrelated events. The diseased individuals with a second disorder, or a complication of the original disease, are more likely to be represented in a hospital-based sample than other members of the general population. The causes of bias include the burden of symptoms, access to care, and popularity of certain institutions (particularly with respect to current practices of admission). Differential rates of admission will be reflected in biased estimates of the relative risks.

This type of bias is more common in observational studies, in particular case-control studies. Since the subjects are randomized after selection, this type of bias is less common among experiments.

Example:

Household interviews were performed on random samples of the general population asking about musculoskeletal and respiratory diseases and recent hospitalizations. In the general population, there appeared to be no association between these two disorders (OR = 1.06), but in the subset of the population who had been in hospital during the previous six months, individuals with musculoskeletal disorders were more likely to have respiratory disease than not (OR = 4.06). This occurred

because individuals with both disorders were more likely to be hospitalized than those with only one of the disorders. This finding is illustrated in Table 6.2.

TABLE 6.2 DISEASES OF THE BONE AND ORGANS OF MOVEMENT WITH AND WITHOUT RESPIRATORY DISEASE

		Diseases of bone and organs of movement					
		General population			Persons hospitalized in previous six months		
		Yes	No	Total	Yes	No	Total
Respiratory disease	Yes	17	207	224	5	15	20
	No	184	2376	2560	18	219	237
	Total	201	2583	2784	23	234	257
Odds ratio		1.06			4.06		

c. *Non-response bias*

This type of bias is due to refusals to participate in a study. The individuals concerned are likely to be different from individuals who do participate. Non-respondents must be compared with respondents with regard to key exposure and outcome variables in order to ascertain the relative degree of non-response bias.

Non-response bias is common in all types of studies, but is more serious in observational studies. In particular, sample surveys are more prone to this type of bias. If the non-response is similar in the exposure and non-exposure groups (or cases and controls), this may not be a serious problem. Sufficient information about related variables should be included in data collection instruments in order that we can verify the effect of non-response bias on the results. Maximizing the response rate in surveys is one way to minimize this type of bias. In randomized controlled trials, it is possible to collect information on related factors that might shed light on the seriousness of the problem by prospectively collecting information and comparing it.

Example:

> In a mailed questionnaire study of the smoking habits of US veterans, it was noted that 85% of non-smokers, but only 67% of cigarette smokers returned the questionnaire within 30 days. Pipe and cigar smokers had an intermediate response rate.

6.2.2 Ascertainment or information bias

Information bias is a distortion in the estimate of effect due to measurement error or misclassification of subjects according to one or more variables. Some specific types of information bias are discussed below.

a. Diagnostic bias

Diagnostic bias may occur due to the performance of a disproportionately high number of diagnostic procedures on cases, as compared with controls. In a cohort study, knowledge of a subject's prior exposure to a possible cause may influence both the intensity and the outcome of the diagnostic process. Knowledge that an individual has worked in the rubber industry, for instance, may lead to a more intensive search for bladder cancer than would occur if the person had worked in another industry.

In a case-control study, if the disease outcome is one with few clinical manifestations, and requires laboratory tests or diagnostic procedures to detect it, the disease may be missed in the control group if they are not adequately examined prior to inclusion in the study. For example, in order to ascertain the presence of endometrial cancer in individuals exposed, or not exposed to estrogen therapy, the same diagnostic procedures must be performed for both groups at the same frequency. This bias can be reduced by having the control group selected from persons who went through the same diagnostic procedures as did the case group, and by using only those with negative results as controls.

Similar bias can also occur in experimental studies, although this is rare, due to the development of, and strict adherence to study protocols that avoid these types of problems.

In general, 'blinding' of persons who are reporting tests, by denying them clinical information about which are cases and which are controls (or to which treatment group they have been allocated), and submitting cases and controls to equally rigorous diagnostic preparation, will help reduce this type of bias.

b. Recall bias

An error of categorization may occur if information on the exposure variable is unknown or inaccurate. Ascertainment of exposure to drugs by history alone, recollection by controls of exposure variable, and a more intense search by investigators for exposure variables among cases, may lead to this type of bias. The recall by both cases and controls may differ in both amount and accuracy. Cases are more likely to recall exposures, especially if there has been recent media exposure on the potential causes of the disease.

Example:

In questioning mothers whose recent pregnancies had ended in fetal death or malformation (cases), and a matched group of mothers whose pregnancies had ended normally (controls), it was found that 28% of the former, but only 20% of the latter reported exposure to drugs. This could not be substantiated either in earlier prospective interviews or in other health records.

This type of bias can be avoided by strict adherence to a developed protocol, administered in a standard fashion by 'blinded' investigators, and by using recorded data to supplement information obtained from records and interviews.

6.3 Effect of selection and ascertainment bias on odds ratios observed in case-control and cohort studies

The biases mentioned in the previous section can alter the odds ratio, and thus potentially lead to an invalid conclusion. The potential effect is illustrated (in general terms) in Table 6.3.

**TABLE 6.3 EFFECT OF BIAS ON ODDS RATIOS OBSERVED IN CASE-CONTROL
AND COHORT STUDIES**

Type of bias	Effect on odds ratio	
	Case-control	Cohort
Selection bias		
Prevalence-incidence	⇑ or ⇓	unlikely
Berkson's bias	⇑ or ⇓	not applicable
Non-response	⇑ or ⇓	⇑ or ⇓
Measurement bias		
Diagnostic	⇑	⇑
Recall	⇑	not applicable

The prevalence-incidence bias can either increase or decrease the odds ratio in a case-control study, but this is unlikely to occur in a cohort study or experiment.   The non-response bias can influence both case-control and cohort studies, as well as experiments, and can occur in either direction.   Selection biases are the most difficult to avoid. The prevalence-incidence bias cannot be prevented in a case-control study, but is at least partially measurable.   The admission-rate bias is neither preventable nor measurable.   Non-response bias can be both prevented and measured.

Of the ascertainment biases, the diagnostic bias will inflate the odds ratio in both case-control and cohort studies.   Recall bias will also inflate the odds ratio in a case-control study, but is not applicable to a cohort study. Both of these biases are preventable.

Selection biases make it impossible to generalize the results to all patients with the disorder of interest, while the measurement biases influence the validity of the study conclusions.

Since biases are difficult to control in most cases, care should be taken to prevent their occurrence by the choice of appropriate design, development of strict protocols and adherence to these protocols.   In the worst case, when these biases cannot be prevented, the potential biases should at least be measured, and possible statistical adjustments of results considered.

6.4   Confounding

Confounding is a special type of bias.   The effect of the factor under consideration is mixed up with effects of other factors not directly relevant to the study question.   An exposure, E is said to be confounded with another factor, C with respect to its effect on a disease, X, if both C and E are associated with the disease, and C and E are associated with each other.   The confounding is manifested in the study results when the factor, C appears unequally among the exposed and unexposed groups;   the comparison of disease incidence or prevalence in the two groups is mixed with the different presence of the factor, C. This is the only type of bias that can often be corrected (if appropriate measures have been taken during the study) by statistical adjustments.

An important consideration when dealing with confounding is that both factors are potential risk factors for the disease;   which one is the cause and which is confounding depends on the study objective. For example, when studying the effect of exposure to asbestos dust (working in asbestos mines) on lung cancer, cigarette smoking is a confounder.   We know that cigarette smoking is closely associated with lung cancer, and that miners tend to smoke more often than non-miners.   On the other hand, if the question of interest was the association of smoking and lung cancer, exposure to asbestos dust could be a confounder.

Confounding is a form of bias, and therefore affects the validity of the study; estimates of the risk coefficients may be systematically higher (or lower) than the true risk.   Adjusting for confounding will improve the validity but reduce the precision of the estimates.   Since it is possible to adjust statistically for confounding, if information on the potentially confounding variables has been collected, there is a tendency to adjust for all potential confounders.   This is counterproductive: one would lose statistical power (precision) and might not gain much in terms of validity if the factors considered were not confounders.   Before adjusting for confounders therefore, both conditions for confounding should be verified.   For a detailed discussion on confounding, see Kleinbaum, Kupper and Morganstern.

When designing a research project, therefore, careful consideration should be given to what are the risk factors of interest, and what could be potential confounders (known risk factors that are of no particular interest in the present study, and that might have an association with the hypothesized risk factors).   Being a type of bias, it is best to avoid the problem if we can, and to collect relevant information if we cannot avoid the problem.

Example:

Suppose that one wants to investigate a postulated causal connection between alcohol consumption and myocardial infarction. Smoking is known to be a cause of this disease; alcohol intake and smoking are known to be correlated. Suppose that alcohol consumption is in fact not a cause of myocardial infarction. By virtue of its association with smoking, however, alcohol intake would be found to be associated with, and apparently to increase the risk of the disease. One might even find an apparent dose-response relationship between alcohol intake and myocardial infarction, since heavy drinkers are often heavy smokers as well. In order to disentangle the effects of smoking and alcohol intake, one may stratify the subjects (both cases and controls) into smoking and non-smoking groups, and within each subgroup, look for an association between alcohol intake and myocardial infarction. Table 6.4 illustrates the effect of confounding in this situation.

**TABLE 6.4   RELATIONSHIP OF ALCOHOL CONSUMPTION
TO MYOCARDIAL INFARCTION (MI)**

A.   Ignoring smoking

Alcohol intake	MI	Control
Yes	71	52
No	29	48
Total	100	100

Odds ratio = 2.26, χ^2 = 7.62, P = 0.006 (2-sided).

B.   By smoking status

Alcohol intake	Non-smokers		Smokers	
	MI	Control	MI	Control
Yes	8	16	63	36
No	22	44	7	4
Total	30	60	70	40
Odds ratio	1.0		1.0	

The statistically significant elevation in risk (OR = 2.26, P<0.01) in the analysis that ignores smoking is spurious. Among non-smokers, the estimated OR for myocardial infarction being associated with alcohol intake is 1.0, with an identical estimate among smokers. The effect in Table A is therefore said to be due to confounding with smoking.

One may regard the subgroup specific ORs in Table 6.4 as representing the effect of alcohol, 'adjusted for smoking' on the risk of myocardial infarction. Conceptually, the effect of smoking has been held constant, although not in an experimental sense. If the two ORs (with and without smoking) were not the same, a pooled estimate of the effect of alcohol intake on MI would not be easy. Statistical methods of adjustment incorporate the use of standardization (using some hypothetical population as standard, so that both exposure and non-exposure groups have a similar distribution of the confounding factor). A common method of such standardization is the Mantel-Haenzel adjusted odds ratio (see Kleinbaum, Kupper and Moganstern for details).

6.5 Options for control of confounding in observational studies

Several methods are available for the control of confounding, either by preventing confounding or by adjusting for it in the analysis.

1. Restriction by study design

This approach to control simply involves specifying narrow ranges of values for one or more extraneous variables in determining admissibility into the study (e.g. restriction to white males only, or to ages between 40 and 50 years). The restriction applies to both index and comparison groups (cases and controls, or exposed and unexposed). This has the effect of removing the confounding variables and retaining a relatively homogeneous group for comparison. The disadvantage of this approach is that the generalizability of the study is limited to the narrow group included in the study. While the study would have external validity to the narrowly defined population, it would not be very useful for the general population of interest.

2. Matching

Matching involves the use of constraints in the selection of the comparison groups, so that the index and comparison groups have similar distribution with respect to the potentially confounding variable.

A common example is when the controls are selected to match the cases for age and gender. By making such a choice, age and gender will no longer be confounding variables (even though they may be associated with the disease, the association of the exposure to the disease is not confounded by these). While this is less restricted than selecting a narrow population of interest, it imposes the restriction that the population of interest is limited to what has been observed in the index group.

Analysis of results from matched studies will need to incorporate the matching design (the two groups are not statistically independent) and often precision is reduced. For example, if 100 cases and 100 controls are used in a matched study, this is like having only 100 observations (100 matched pairs) and the statistical power is approximately 60% compared with the unmatched study with 100 cases and 100 controls. Therefore, matching has to be done judiciously. In addition, matching for several variables simultaneously can lead to serious 'overmatching' in that any potential association gets washed out, and results are never statistically significant.

3. Stratification in the analysis without matching

This option essentially involves restriction of the analysis (rather than the sampling scheme) to narrow ranges (strata) of the extraneous variable. Pooling of the results from the various strata may be possible, if there is no interaction between the two factors. An example was presented in Table 6.4.

4. Mathematical modelling in the analysis

This approach involves the use of advanced statistical methods of analysis, such as multiple linear regression, logistic regression, etc. This is a form of stratification in the analysis and pooling of the information, except that the stratification and pooling is done under the assumption of some mathematical form of relationship. Specific types of relationships may be explored by these methods, and these can be statistically more powerful than the individual stratified analysis. For more details, see Kleinbaum, Kupper and Morganstern, or Hosmer and Lemeshow.

6.6   Recommendations for minimizing bias in analytical studies

1.   Cases should be limited to incident cases, and should be chosen as homogeneous entities or as random samples of all cases.

2.   Definitions, ascertainments and exclusions must always be made explicit, and this should be done in advance.

3.   At least two control groups should be chosen:

   a.   a hospital-based group, preferably from among patients who have undergone the same diagnostic procedures as the cases; controls may either be matched to the cases, preferably on a stratified basis, or chosen as a random sample of potential controls;

   b.   a community-based control group.

4.   Analysis should be complete. All known potential confounders, if not already considered in the matching process, should be the subject of analysis by stratification or multivariate techniques.

6.7   References and further reading

Fletcher H.R., Fletcher W.S., Wagner H.E. Clinical epidemiology: the essentials. London, Williams and Wilkins, 1982.

Hosmer D.W., Lemeshow S. Applied logistic regression. New York, John Wiley and Sons, 1989.

Kleinbaum G.D., Kupper L.L., Morganstern H. Epidemiologic research: principles and quantitative methods. New York, Van Nostrand Reinhold, 1982.

Sackett D.L., Haynes R.B., Guyatt G.H., Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston, Little, Brown and Company, 1991.

Chapter 7

Basic Risk Measurement

7.1   Introduction

Health research involves the estimation of risk of a disease, or the probability of an outcome in one form or another. Measures of risk vary according to the type of variable and the design of the study. In this chapter, we will review some commonly used indicators of risk and their interrelationships.

7.1.1 Review of probability notations

Risk is essentially a measure of probability. Almost all aspects of statistical analysis of epidemiological data can be viewed in terms of probability concepts, and the conclusions are almost always accompanied by the calculation of probabilities of various events. In hypothesis testing, the uncertainty in research conclusions is related to the probability of the data being concordant with the hypothesis. In descriptive studies, the uncertainties in the estimated parameters are expressed by confidence intervals, using measures of probability.

Probability is a measure of uncertainty. It is often expressed as a relative frequency: of all the possible events, how likely is it that the event under consideration will occur? Obviously, probability refers to future events, but it is estimated from our past experience. For example, in a clinical trial, if 60% of the people taking a medication find relief from their symptoms, one estimates that the probability of symptom relief in future is 60% (for every 100 patients with the condition taking this medication, 60 will show relief). It should also

be noted that probability is in fact a measure for a group of people, and the concept as it applies to individuals, is difficult to explain. Yet, we all understand what is meant when we say that the chance of symptom relief is 60%.

Probability is a number, by definition, between 0 and 1 (0% and 100%). Since it is the relative frequency, the numerator is the number of events, and the denominator represents all the people at risk of the event. For example, if N_D persons in a population of N have a disease, the probability that a randomly chosen person from this population will have the disease is N_D/N.

If, in the same population, N_S denotes the number of people exposed to a factor, and N_{SD} denotes the number of people among those with the factor who have the disease, then the conditional probability of disease, given the presence of the factor, is N_{SD}/N_S. The conditional probability is written as $P(D|S)$, and one can see that:

$$P(D|S) = \frac{N_{SD}}{N_S} = \frac{(N_{SD}/N)}{(N_S/N)} = \frac{P(D \text{ and } S)}{P(S)}$$

Three basic laws of probability are used in calculations of probability:

1. Probability is between 0 and 1;

2. $P(A \text{ or } B) = P(A) + P(B)$
    if A and B are mutually exclusive
    $= P(A) + P(B) - P(A \text{ and } B)$
    if they are not;

3. $P(A \text{ and } B) = P(A) \times P(B)$
    if they are 'independent'
    $= P(A) \times P(B|A) = P(B) \times P(A|B)$
    if they are not.

Repeated use of the above rules allows one to compute probabilities for various events in epidemiological studies, and to calculate the appropriate risk estimates and their standard errors.

7.1.2 Use of probability in diagnostic tools

The accuracy of diagnostic tools is often measured by how often the diagnosis correctly identifies diseased people (sensitivity of the test) and how often the tool identifies those who are well (specificity of the test). From these measures, the predictive value of a positive or negative test can be derived.

7.2   Use of probability to assess risks in epidemiological studies

7.2.1 Incidence and prevalence

The basic measure of risk is the probability of disease (or any outcome of interest). Two measures are commonly used: prevalence and incidence. Prevalence measures the probability of having a disease, whereas incidence measures the probability of getting a disease. These can be expressed formally as:

Point prevalence =no. of people with the disease (outcome) in a population at a specific point in time / total population at risk at that time

Period prevalence = no. of people with the disease in a population during a specific period / total population at risk during that period.

In both cases, the numerator is the number of existing cases. This is the measure that is available in a cross-sectional study.

Incidence, on the other hand, has the number of new cases in the numerator. As we saw earlier, there are two ways of measuring incidence, depending on what denominator is used: the cumulative incidence and the incidence density. Both provide estimates of probabilities of acquiring the disease, but the unit of measurement is different in the two methods. Cumulative incidence estimates the probability of acquiring the disease per person, and the incidence density is the estimate of probability of acquiring the disease per person-time.

7.2.2 Measures (indices) of association

Measures of association of an exposure with an outcome always involve the probabilities of the various events. The actual measure to be used depends on the design strategy. Probability can be calculated only when the 'population at risk' can be ascertained.

<u>Cohort studies</u>

When the probabilities of disease in two groups are compared, as in the case of a cohort study, where the probability of disease among the exposed group is compared with the probability of disease among the unexposed, a relative measure is used. The ratio of the two probabilities is called the relative risk (RR).

RR = incidence among exposed / incidence among the unexposed

Either of the two measures of incidence may be used. For example, suppose a cohort study of 400 smokers and 600 non-smokers documented the incidence of hypertension over a period of 10 years. The following table summarizes the data at the end of the study:

		Hypertension		
		Yes	No	Total
Smoking	Yes	120	280	400
	No	30	570	600
	Total	*	*	1000

The probability of hypertension among smokers, P(H|S), also denoted as I_e (incidence among exposed)

= 120/400 = 0.30.

The probability of hypertension among non-smokers, P(H|nonS), also denoted as I_o (incidence among unexposed)

= 30/600 = 0.05.

RR = I_e / I_o = 0.3/0.05 = 6.0.

An RR of more than 1 indicates the factor to be positively associated with the disease (exposure increases the chance of the disease, e.g. smoking) and an RR of less than 1 indicates a protective factor (exposure decreases the chance of disease, e.g. vaccination). Notice that, in the cohort study, since the 'population at risk' is followed, and all new cases of disease (within the specified time period) have

been identified, a true measure of probability (risk) can be computed. The totals down the column (total no. of diseased and non-diseased persons) are not proper denominators for any probability.

Another measure that is commonly derived from the probabilities of disease in the two groups, is the attributable risk (AR): the excess risk for the exposed group compared with the unexposed group. This is simply the difference between the two probabilities:

$$AR = I_e - I_o = 0.30 - 0.05 = 0.25$$

Twenty-five per cent of the new cases of hypertension among the exposed group can be attributed to smoking. Sometimes the AR is expressed as a percentage of the incidence in the total population, and this is called the attributable risk percent (AR_P):

$$
\begin{aligned}
AR_P &= (I_T - I_o / I_T) \times 100 \\
&= [(0.15 - 0.05) / 0.15] \times 100 \\
&= 66.6\%
\end{aligned}
$$

This is mathematically equivalent to:

$$AR_P = P_e (RR-1) / [1 + P_e (RR-1)]$$

where P_e is the proportion of the population exposed to the factor.

Another measure of attributable risk is the etiologic (attributable) fraction (EF), which is the ratio of the AR over the incidence in the exposed group. This measure answers the question, 'what proportion of cases among the exposed group can be attributed to the exposure?'

$$
\begin{aligned}
EF &= (I_e - I_o) / I_e \\
&= (0.30 - 0.05) / 0.30 \\
&= 83.3\%
\end{aligned}
$$

Mathematically, this is equivalent to:

$$
\begin{aligned}
EF &= (RR - 1) / RR \\
&= 1 - 1/RR
\end{aligned}
$$

Thus, in the above study, 66.6% of the hypertension in the population can be attributed to smoking, and 83.3% of hypertension among smokers can be attributed to smoking. Both of these measures can be used to estimate the number of cases that could be prevented if the risk factor were removed from the population, and therefore is a useful public health tool when developing programmes for the prevention of diseases.

In general, in a cohort study, the results are tabulated to provide the 2x2 table of exposure and disease status, as follows:

		Disease (outcome)		
		Present	Absent	Total
Exposure	Present	a	b	a+b
	Absent	c	d	c+d
	Total	a+c	b+d	N = a+b+c+d

Then,

$$I_e = a/(a+b); \quad I_o = c/(c+d)$$

$$RR = a(c+d) / c/(a+b)$$

$$AR = [a/(a+b)] - [c/(c+d)]$$

$$AR_P = [(a+b/N) \ (RR-1)] / [1+(a+b/N) \ (RR-1)]$$

$$EF = 1 - 1/RR.$$

Case-control studies

In case-control studies, clearly the incidence is not measurable, and hence the relative risk is not estimable. However, if the disease is rare, an approximation can be made. This measure is called the odds ratio (OR). Assume that the above table represents a case-control study. Now the totals (a+c) and (b+d) are obtained by design, and hence are valid denominators, while the numbers (a+b) and (c+d) are the results of the study, and not valid denominators. The only measurable probabilities are the prevalence of risk factors among the

diseased (a+c) and non-diseased (b+d) groups.   A measure related to probability constitutes the 'odds' of an event.   Among the diseased, the odds of being exposed are a/b, while among the non-diseased, the odds are c/d. The ratio of these odds is called the odds ratio.

Notice that OR = ad/bc is easily calculated from the 2x2 table of results.   It can also be shown that this is a good approximation to the RR, when the disease is rare.

$$RR = [a/(a+b)] / [c/(c+d)] = (ac+ad)/(ac+bc).$$

If the disease is rare, (ac) is much smaller than (ad) and (bc), and the above becomes approximately equal to ad/bc = OR.

Thus, in a case-control study, the odds ratio (OR) is used as a measure of association of the disease and the risk factor.   Notice that the attributable measures are not possible in this situation, since the cases and controls are preselected, and hence the incidence cannot be calculated.   However, some researchers use OR estimates to substitute for RR in the equations for AR and EF, to obtain an equivalent measure; in general this is not justified.   In the special case where the prevalence of the disease is very low, arguments for the case of RR being approximately equal to OR when incidence is low, may be possible.

In our example, if the same results were obtained from a case-control study of 150 cases of hypertension and 850 people without hypertension, the table would appear as follows:

		Hypertension		
		Yes	No	Total
Smoking	Yes	120	280	*
	No	30	570	*
	Total	150	850	1000

OR = 120x570 / 30x180 = 8.14.

Cross-sectional studies

In the case of cross-sectional studies, the population sampled is the total population. Therefore, both the prevalence of disease and the prevalence of the risk factor can be estimated. Here, all the elements of the 2x2 table are valid measurements, and allow one to calculate the appropriate probabilities. Note, however, that the probabilities are not 'risk of acquiring the disease', but rather the prevalence measure.

All the measures stated above can be computed from the 2x2 table and, apart from the fact that we are talking about prevalence and not incidence, the explanations are valid. The RR and OR would be calculated in the same way, and other quantities such as the AR and EF can also be calculated. If the prevalence and incidence are similar, these measures may have the same interpretations. More importantly, testing of hypotheses regarding the various probabilities would be valid in this type of design, and would provide the basis for further refinement of the risk estimates in studies with better designs (cohort, quasi-experimental or experimental).

In the table of observations, all the cells would now have valid numbers. The above table, if it had arisen from a cross-sectional study, would appear as:

		Hypertension		
		Yes	No	Total
Smoking	Yes	120	280	400
	No	30	570	600
	Total	150	850	1000

7.2.3 Statistical variation in the measures

Notice that all the above measures are point estimates of the appropriate measure of association. Since the studies are usually based on samples, we need to identify the 'random error' associated with these estimates, i.e. what is the possible range of values within

which the true measure lies. One needs to develop the probability distribution of these measures, and from this, calculate an appropriate confidence interval. These concepts are discussed in more detail in Chapter xx. Suffice it to say that we can usually calculate a 'standard error' of the estimates and, using this, obtain the confidence intervals using the normal approximation (i.e. estimate +/- 2 standard error would give approximately a 95% confidence interval for the risk measure).

7.3   References and further reading

Abramson J.H. Survey methods in community medicine, 2 ed. New York, Churchill Livingstone, 1979.

Barker D.J.P., Bennet F.J. Practical epidemiology. New York, Churchill Livingstone, 1976.

Beaglehole R., Bonita R., Kjellstrom T. Basic epidemiology. Geneva, WHO, 1993.

Fleiss J.L. Statistical methods for rates and proportions. New York, John Wiley and Sons, 1981.

Kelsey J.L., Thompson W.D., Evans A.S. Methods in observational epidemiology. Oxford, Oxford University Press, 1986.

Kleinbaum D.G., Kupper L.L., Morganstern H. Epidemiologic research. New York, Van Nostrand Reinhold, 1982.

Lillienfeld A.M., Lillienfeld D.E. Foundations of epidemiology, 2 ed. Oxford, Oxford University Press, 1980.

Chapter 8

Statistical Analysis of Data

8.1 Introduction

In most research studies, the information collected represents only a sample from the population of interest (target population). Drawing conclusions about the population, whether it is a simple descriptive study or a randomized controlled trial, depends on statistical analysis of the data. This manual is intended to assist in the preparation of a research proposal, and not with data analysis. However, since the choice of the design has a direct impact on the analysis of the data, it is important to have an idea of the type of analysis anticipated when designing the study. Therefore, we will briefly review the important aspects of statistical analysis.

8.2 Basis for statistical analysis

The fundamental principles of probability theory (briefly reviewed in Chapter 7) are used in statistical inference. All the inferences are based on three primary entities: the population (U) that is of interest, the set of characteristics (variables) of the units of this population (V), and the probability distribution (P) of these characteristics in the population.

The population (U)

The population is a collection of units of observation that are of interest, and is the target of the investigation. For example, in determining the effectiveness of a particular drug for a disease, the population would consist of all possible patients with this disease. In

determining the prevalence of the incidence of HIV infection among commercial sex workers in a country, the population would consist of all the commercial sex workers in the community. The 'population' here is synonymous with the 'target population' identified in Chapter 7.

It is essential, in any research study, to identify the population clearly and precisely. The success of the investigation will depend to a large extent on the identification of the population of interest. Often, the population of interest is not observable, and a smaller population is identified as the subject of investigation. For example, in clinical trials, some patients are excluded for various reasons prior to randomization, and the studied population is therefore somewhat different from the target population. This distinction should be clear at the beginning of the study, but also at the time of data analysis and interpretation, so that the inferences drawn from the study will be valid.

The variables (V)

Once the population is identified, we should clearly define what characteristics of the units of this population (subjects of the study) we are planning to investigate. For example, in the case of the HIV study above, one needs to define HIV (reliable and valid method of identifying HIV in people), and what other characteristics of the people (e.g. age, sex, education, etc.) one intends to study. Clear and precise definitions and methods for measuring these characteristics (a simple observation, a laboratory measurement, or a battery of tests using a questionnaire) are essential for the success of the research study.

The variables are characterized in many ways; for statistical considerations, the variables are usually classified as discrete or continuous. Discrete variables are those in which only a small number of values is possible (e.g. sex: male, female), incidence of a disease (yes, no)). Continuous variables are those which, theoretically, can take any value within a specified range of minimum and maximum value (e.g. age, blood pressure). There are some variables that are discrete in nature, but the number of categories make them similar to continuous variables, and these are considered as continuous in most statistical calculations (e.g. number of years of schooling, number of people in a household).

The probability distribution (**P**)

The most crucial link between the population and its characteristics, which allows us to draw inferences on the population based on sample observations, depends on this probability distribution. The probability distribution is a way to enumerate the different values the variable can have, and how frequently each value appears in the population. The actual frequency distribution is approximated to a theoretical curve that is used as the probability distribution.

Common examples of probability distributions are the binomial, Poisson and normal. Most statistical analyses in health research use one of these three common probability distributions. For example, the incidence of a relatively common illness may be approximated by a binomial distribution, whereas the incidence of a rare condition (e.g. number of deaths from motor vehicle accidents) may be considered to have a Poisson distribution. Distributions of continuous variables (blood pressure, heart rate) are often considered to be normally distributed.

Probability distributions are characterized by 'parameters': quantities that allow us to calculate probabilities of various events concerning the variable, or that allow us to determine the value of probability for a particular value. For example, the binomial distribution has two parameters: n and π. The binomial distribution occurs when a fixed number (n) of subjects is observed, the characteristic is dichotomous in nature (only two possible values), and each subject has the same probability (π) of having one value and ($1-\pi$) of the other value. The statistical inference then involves finding out the value of π in the population, based on an observation of a carefully selected sample.

The normal distribution, on the other hand, is a mathematical curve represented by two quantities, μ and σ. The former represents the mean of the values of the variables, and the latter, the standard deviation. (Definitions in section 8.3.3.)

The type of statistical analysis done depends very much on the design of the study: in particular, whether the study was descriptive, and what sampling design was used to draw the sample from the population.

8.3   Descriptive studies

In descriptive studies, the object is to estimate the values of the parameters of the probability distribution, or a function of these parameters. Based on what was observed in the sample, an estimate (best guess) of the values in the population is made, and a measure of the accuracy of this estimate is obtained. The measure of accuracy is based on what is known as the sampling distribution of the estimate.

8.3.1 Accuracy of estimates

When a descriptive study is conducted and an estimate (E) of a parameter is obtained from the study, we need to know how this value, E would change if we took another sample. The distribution of values of E over different repetitions of sampling (under identical conditions to the ones we have already employed) is known as the sampling distribution of E. The sampling distribution can be empirically determined by actually repeating the process. Clearly, this is both difficult and unwarranted. It is possible to get an approximate idea of the sampling distribution, purely based on sampling theory.

Once the sampling distribution is obtained, we can answer questions such as 'how close is my estimate likely to be to the true value of the parameter?' Obviously, we cannot get a 100% certain answer to this question, because we have only observed a sample. However, based on the sampling distribution, we can state with a certain amount of confidence (e.g. 95% sure) that it will be within ±x of the true value. This interval is known as the confidence interval. The greater the confidence in the statement, the larger is the value of x (wider interval). As we see below for specific examples, it is also known that the width of the interval for the same amount of confidence will decrease with an increase in sample size. Intuitively, the more information we have (large n) the more confident we are (smaller width of interval, or larger confidence for same interval).

8.3.2 Estimation of parameters of the binomial distribution

When the study deals with a dichotomous event (such as incidence of a disease), the objective is to obtain an estimate for the probability of the event (incidence rate) occurring in the population. Based on the binomial probability distribution, it has been shown that the best estimate is the sample proportion, p (number of events in the sample/ sample size, n).

In order to assess how accurate this estimate is (how close is p to the true value, π), we need to know how much variability is expected in p in repeated samples using the same design (sampling distribution of p). It has been shown that, for p, the distribution is approximately normal, with mean p and standard deviation, and $s = \pi(1-\pi)/n$ (s is known as the standard error of p). Using the properties of the normal distribution, we can then say that the true value of π is within $p \pm 1.96s$, with 95% confidence.

Example 1

In a study to determine the prevalence of HIV infection among commercial sex workers (CSW), a sample of 150 CSW was tested, and 42 were found to be positive for HIV. The estimate for HIV prevalence was therefore 28%, with a standard error of 3.67%. The 95% confidence interval for HIV prevalence among CSW in this community is therefore 28±1.96*3.67 = (20.82%, 35.18%), i.e. based on this survey, we can state with 95% confidence that the true prevalence could be as low as 21%, or as high as 35%.

Notice that, in Chapter 7, we discussed many parameters or functions of parameters from binomial distributions when we discussed the incidence and prevalence and the risk ratios.

The RR and OR that we get from cohort and case-control studies are the estimates of true risk ratios in the population from which the study samples were selected. To complete the picture, therefore, we need to calculate the sampling distributions of these estimates. In most cases, the sampling distributions are assumed to be approximately normally distributed (a statistically acceptable result if the sample size is large and the sampling is done using probability methods), so that we need only to calculate the standard error of these estimates to construct confidence intervals. Most computer programs that calculate relative risks or odds ratios will also report their standard errors, and in some cases the confidence intervals.

8.3.3 Estimation of parameters for normal distribution

For a variable, X that has a normal distribution, we would need to know the mean μ and the standard deviation, σ. The best of these parameters is the sample mean $\bar{x}$ (arithmetic average of all the observations in the sample) and the sample standard deviation,

$$s = \sqrt{\sum_i (x_i - \bar{x})^2 / (n\text{-}1)}$$

[The Normal distribution has the property that it is a symmetric probability distribution, the centre of the distribution is μ.

Also, the mean $\pm$ 1.96 (standard deviation) contains 95% of the values of the variable (i.e. the probability that the variable has values within this interval is 95%).]

Another reason that the normal distribution is commonly used in statistical inference is that most sample functions (sample mean, risk ratios, correlation coefficient, etc.) have the normal distribution as the sampling distribution, if the sample size is sufficiently large.

Most of the inferences in health research involve only inferences on the mean value. The sample mean has a normal distribution with mean m and standard deviation (standard error of the mean), $s/\sqrt{n}$. Thus, the 95% confidence interval for the population mean, m is therefore:

$$\bar{x} \pm 1.96 s / \sqrt{n}$$

Or, more simply, sample mean $\pm$ 2* (standard error of mean). For a more detailed description of common estimation problems and formulae for confidence intervals, see Kleinbaum, Kupper and Morgenstern, or Glantz.

8.4 Analytical studies

In contrast to descriptive studies, analytical studies involve the testing of hypothesis in addition to description of the population. The study will have formulated research hypotheses, and on the basis of

the observations in the research study, we need to draw conclusions as to the validity of these hypotheses. The inference is therefore a two-step process: estimate the parameters of the relevant probability distributions; test hypotheses (also known as testing of significance) involving these parameters.

8.4.1 Statistical tests of hypotheses

A test of hypothesis has several steps:

<u>Step 0. Identify the null hypothesis</u>

This is a re-statement of the research hypothesis in the 'null' form, i.e. 'no effect of treatment', 'no difference in survival rates', 'no difference in prevalence rates', 'relative risk is one', etc. The null hypothesis is often stated with the research objectives. The null hypothesis should be 'testable', i.e. it should be possible to identify which parameters need to be estimated, and it should be possible to estimate the parameter, its standard error and the sampling distribution, given the study design.

<u>Step 1. Determine the levels, α and β of errors acceptable in the inference</u>

Since the inference is based on a sample of the population, one will never be absolutely sure if the hypothesis is true or not in the population. The decision is a dichotomous one: to accept the null hypothesis H, or to reject H_0. Two types of errors in inference are possible. The type I error (α) is the probability of falsely rejecting the null hypothesis, and the type II error (β) is the probability of falsely accepting the null hypothesis. These are summarized in the table below:

	'Truth' (in the population)	
Decision (based on sample results)	H_0 is true	H_0 is false
Accept H_0	No error	Type II or β
Reject H_0	Type I or α	No error

Notice that the aim of the research study is to minimize both α and β; however, they work in opposite directions. If we decrease one, the other tends to increase. The researcher often designs the study to achieve a desired level for α, and minimize β for this situation. The statistical testing of hypothesis, therefore, is often done with a choice of α and the best statistical test available that will minimize β. The choice of α and β is made after determining the consequences of each of the errors, and is fixed at the time of design.

Step 2. Determine the best statistical test for the stated null hypothesis

This depends on the design, the type of variables, and the type of the probability distribution of the variable. For example, suppose that the null hypothesis is that the prevalence rates of a disease among two population groups are the same, and simple random samples have been obtained from the two population groups independently (design). The variable is the disease, which is a (discrete) dichotomous variable, and the sample size is fixed. Therefore, a binomial distribution is the probability distribution under consideration, and the prevalence rate is the parameter of the distribution, which is estimated by the sample prevalence rates. These have approximately normal distributions (sampling distribution). Therefore, a z-test or chi-square (χ^2) test (see below) is the most appropriate.

Step 3. Perform the statistical test

This involves calculating the appropriate test statistic (the z or χ^2) and comparing the computed value with its theoretical distribution. If the observed value is outside the limits in which the probability is $<\alpha$ for the sampling distribution, the null hypothesis is rejected.

Step 4. Calculate the power of the test

If the null hypothesis is not rejected, i.e. the computed value of the test statistic is within the limits for the α, then the statistical power of the test $(1-\beta)$ should be computed for some acceptable minimum departure from the null hypothesis. If the power is too low, one would recommend that the study be repeated with a larger sample size. If the power is acceptable, one accepts the null hypothesis.

Sometimes, instead of deciding on 'acceptance' or 'rejection' of H_0, the test statistic is compared with the sampling distribution, and the value of α at which the test would reject the H_0 is calculated. This is called the P-value for the test.

In the above example, if the computed value of z were less than -1.96, or greater than 1.96, or equivalently, if the χ^2 value were above 3.84, one would reject the null hypothesis, with $\alpha = 0.05$.

It should also be noted that rejecting a null hypothesis does not necessarily mean that the effect or difference (departure from the null hypothesis) is 'clinically' significant. The differences may be trivial in terms of practical usefulness, and yet statistically significant, if the sample size is large. For example, an odds ratio of 1.1 can be statistically significant at 5% level of significance, if the sample size is very large (say 100 000), but one would not worry too much about an increase in relative risk by such a small amount. (Of course, it depends on the particular disease, and the smallest difference that makes a significant impact is often called the minimally acceptable difference, and is used in calculating the sample size when designing the study; see Chapter 5)

When we reject a null hypothesis, we usually accept an alternative hypothesis, H_1, which in most cases is the opposite of H_0. For example, if H_0 = the means of two populations are equal, then H_1 = the two means are not equal. This type of alternative hypothesis is called a two-sided alternative. When the mean of one population is too large or too small compared with the other, we reject the null hypothesis. There may be cases in which we are interested only in detecting whether the difference is on one side of the hypothesis (e.g. does the drug improve the survival rate?) In this case, the testing can be one-sided, and the H_0 rejected when the difference is too large and showing the benefit of the drug, but not if the difference is too large and showing that the drug is detrimental. Obviously, since we reject H_0 only half the time, the type I error is reduced; equivalently, for the same type I error, H_0 is rejected more often, increasing the power of the test. The decision to use a one-sided or two-sided test should be made in advance (before data collection), and should be based on solid scientific reasoning, lest the comparison be biased.

8.4.2 Some common statistical tests of hypotheses

Comparison of two proportions (z-test; χ^2 test)

A common test of significance in epidemiological studies involves the comparison of two proportions. Examples include the comparison of incidence rates (in cohort studies) and the comparison of prevalence rates (in case-control or cross-sectional studies). Comparison of proportions involves the testing of a null hypothesis of the form $H_0 : \pi_1 = \pi_2$, where π_1 and π_2 are the probabilities of an event in two independent populations. The common design involves a simple random sample of subjects, taken from the two populations independently, or using some form of matching (e.g. paired observations, such as matched case-control studies with exact matching on age). The event or characteristic, such as the incidence or prevalence of a disease, exposure to a risk factor, belonging to a particular race, etc., is either dichotomous, or is made dichotomous by grouping all the events not of interest into one group (e.g. in a multiracial country such as Canada, the interest may be to compare the white population with the rest). The probability distribution assumed is binomial.

The test of the hypothesis is based on the observed proportions, p_1 and p_2 in the two samples. If H_0 is true, one would expect $(p_1 - p_2)$ to be zero. The sampling distribution of $(p_1 - p_2)$ is approximately normal, with mean $(p_1 - p_2)$ and standard deviation (standard error of the difference) given by the formula:

$$\sqrt{[\pi_1(1-\pi_1)/n_1]+[\pi_2(1-\pi_2)/n_2]}$$

Therefore the test statistic,

$$z = (p_1 - p_2)/\sqrt{p_1(1-p_1)/n_1]+[p_2(1-p_2)/n_2]}$$

has a normal distribution, with mean 0 and standard deviation 1, if the H_0 is true. Under the null hypothesis, $\pi_1 = \pi_2 = \pi$; therefore, the standard error is:

$$\sqrt{[\pi_1(1-\pi_1)(1/n + 1/n_2)]}$$

and is estimated by

$$\sqrt{p(1-p)(1/n_1+1/n_2)}$$

where $p = (n_1 p_1 + n_2 p_2)/(n_1 + n_2)$.

Equivalently, $\chi^2 = z^2$ has a chi-square distribution with one degree of freedom. The statistical test therefore, is to calculate z or χ^2 and compare with the appropriate distribution. For example, if a = 0.05, the cut point for z is ± 1.96, and the cut point for χ^2 is 3.84. Notice that χ^2 can also be calculated in a simple way from the two-way table, as illustrated below:

OBSERVED FREQUENCIES

		Population 1	Population 2	Total
Event	Yes	O_{11}	O_{21}	a
	No	$O_{12} = n_1 - O_{11}$	$O_{22} = n_2 - O_{21}$	b
	Total	n_1	n_2	$n = n_1 + n_2 = a + b$

If H_0 is true, based on the sample sizes, we have the following:

EXPECTED FREQUENCIES

		Population 1	Population 2	
Event	Yes	$E_{11} = a^*n_1/n$	$E_{21} = a^*n_2/n$	a
	No	$E_{12} = n_1 - E_{11} = b^*n_1/n$	$E_{22} = n_2 - E_{21} = b^*n_2/n$	b
	Total	n_1	n_2	n

$$\chi^2 = \sum (O - E)^2 / E$$

Where the summation is over the four cells of the 2x2 table, O = observed frequency and E = expected frequency.

<u>Example 2</u>

In a cohort study of low birth weight, 250 women of Chinese origin and 150 women of Indian origin were followed throughout their pregnancies for various risk factors for low birth weight (birth weight less than 2500 grams). Twelve Chinese women and 18 Indian women gave birth to infants weighing less than 2500 grams. The research question was whether the incidence of low birth weight was higher among the Indian women.

Variable: low birth weight (dichotomous: yes/no).

Parameter of the binomial distribution π = incidence rate.

Null hypothesis: $\pi_C = \pi_I$: type I error = 0.05.

Data:

	Chinese	Indian
Low birth weight	12	18
Normal birth weight	238	132

Estimates of incidence rates:

Chinese: p_C = 12/250 = 4.8%

Indian: p_I = 18/150 = 12%.

Test procedure:

(a)   z = (4.8 - 12)/ $\sqrt{(4.8 * 95.2/250) + (12 * 88/150)}$

= (-7.2/2.98) = -2.42

cut point for z = ±1.96

Since the calculated z is less than -1.96, we reject the null hypothesis, and conclude that the incidence rates are different in the two populations. [The difference in the incidence rates is statistically significant (P<0.05).]

(b) Expected frequencies for the four cells of the above table are:

	Chinese	Indian
Low birth weight	18.75	11.25
Normal birth weight	231.25	138.75

$$\chi^2 = [(-6.75)^2/18.75 + (6.75)^2/11.25 + (6.75)^2/231.25 + (-6.75)^2/138.75)] = 7.01,$$

which is larger than the chi-square, with one degree of freedom cut point for a 5% tail area. Therefore, we reject the null hypothesis at the 5% level of significance, and conclude that the two incidence rates are different.

Comparison of incidence in cohort studies, or prevalence in case-control studies

A specific example of the comparison of incidence is a cohort study. In such a case, the index of comparison might be the relative risk, rather than the risk difference as above. The null hypothesis, $I_1 = I_2$ may be re-stated to the null hypothesis, RR = 1. We need to find the sampling distribution of the sample risk ratio, rr, in order to test this hypothesis. Since it is a ratio, the function, ln(rr) is assumed to have a normal distribution with mean zero. Based on this, a test of significance involves the computation of the standard error of ln(rr), using the test statistics, z = ln(rr) / s.e.(ln(rr)) as above. In practice, however, the test of significance is done on the hypothesis of equal incidence rates, and the chi-square test is appropriate. For further discussion, see Kleinbaum, Kupper and Morganstern.

Comparison of two proportions when the samples are matched

When the two samples are matched, especially in a one-to-one matching, the resulting observations are not statistically independent. Therefore, the standard error of the difference will involve a covariance term. Moreover, the difference may not have a normal distribution. Therefore, the z test for two independent samples is no longer valid. There are statistical tests that take into account the dependency between the samples. One test in particular, the McNemar's chi-square, is worth mentioning. Suppose the two

samples are matched one-to-one, so that we have n pairs of observation (total 2n observations). McNemar's test involves separating these n pairs into concordant (both members of the pair have the event, or neither has) and discordant (one member of the pair has the event, the other does not). Characterizing the event as + and the non-event as -, the four categories of observations and their frequencies are summarized below:

+/+   a;   +/-   b;   -/+   c;   -/-   d

The two concordant groups, +/+ and -/-, are discarded, as they do not provide information on the null hypothesis of equal probability in the two populations. If the null hypothesis were true, one would expect the discordant pairs +/- and -/+ to have equal frequencies, so the expected numbers in these groups are (b+c)/2.

A chi-square based on these sets of two observed frequencies (b,c) and two expected frequencies [(b+c)/2, (b+c)/2] follows a chi-squared distribution with one degree of freedom, if the null hypothesis is true.

The McNemar's $\chi^2 = (r-c)^2 / (r+c)$, and should be compared with a chi-square distribution with one degree of freedom for the test of significance of the null hypothesis. If the numbers of pairs are small, a continuity correction is often applied to this formula:

$$\chi^2 = (r-c-1)^2 / (r+c)$$

Example 3

In a case-control study of nasopharyngeal carcinoma (NPC), 200 cases of NPC were matched to 200 control subjects (patients from the same hospital admitted with other conditions, matched for age, sex and race of the patient). One of the risk factors considered in the study was exposure to Epstein-Barr virus (EBV). The following table summarizes the results for the 200 pairs of subjects in the study, with respect to this risk factor:

No. of pairs	EBV exposure	
	Among cases	Among controls
45	+	+
28	-	+
56	+	-
71	-	-

The null hypothesis is that there is no association between the exposure and the disease, which translates, as the frequencies of the two discordant pairs are equal.

Discarding the 'tied' pairs (+,+), (-,-), the discrepant pairs have frequencies of 28 and 56 respectively. The McNemar's chi-square is therefore $(28-56)^2 / (28+56) = 10.6$. If we choose the type I error $(\alpha) = 0.05$, the cut point for χ^2 is 3.84, and we reject the null hypothesis of no association between the exposure (EBV) and the disease (NPC).

Comparison of two proportions when the sample size is small

All the above tests are based on normal approximation of the test statistics, which depend on the sample size being large. The requirement is that np is more than 5 (the expected frequency in each cell in the contingency table is above 5). When the sample size is too small to have this requirement, the normal approximation may be incorrect. Sometimes a continuity correction is applied to the chi-square, although this is not widely accepted. A test that does not use the normal approximation, the Fisher exact test, is used in such situations. See Glantz for further details.

Comparison of two means (independent samples)

When the variable of interest is a continuous one, the relevant probability distribution is the normal distribution. In such cases, the null hypothesis often takes the form, $H_0: m_1 = m_2$, where m_1 and m_2 are the means of the variable in the two populations, respectively. The test of the hypothesis follows the same steps as in the case of testing the difference in proportions, except that the parameter of

interest is the difference in means, $\bar{x}$. The best estimate of the population mean, m, is the sample mean. Therefore, to test the null hypothesis, we compute the standardized difference in means.

In the case of the two samples being obtained independently (for example, in a clinical trial where the patients have been randomly allocated to two groups, or in an unmatched case-control study), this value has a normal distribution, with mean 0 and standard deviation 1, if the null hypothesis is true. Here, we are assuming that the standard deviations in the two populations, σ_1, σ_2 are known. In practice, however, we seldom know these quantities and they have to be estimated by their respective sample standard deviations. Commonly, it is assumed that the two populations have the same standard deviations, and a pooled estimate of the common standard deviation, s, is used in the calculations:

$$s = \sqrt{[(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}]/(n_1 + n_2 - 2)}$$

Then the standardized difference,

$$t = (\bar{x}_1 - \bar{x}_2)/s\sqrt{(1/n_1 + 1/n_2)}$$

has a student's t-distribution with (n_1+n_2-2) degrees of freedom, if the null hypothesis is true. Therefore, the test would be to compare the computed value of t with table values for the appropriate t-distribution for the chosen α.

Example 4

From a study of the incidence of low birth weight (birth weights of 2500 grams or less) among various ethnic groups in Malaysia, the average birth weights, along with the standard deviations, are given below:

Ethnic group	Mean	N	Std. deviation
Malay	2816.71	458	597.52
Chinese	2692.05	156	577.95
Indian	2914.26	135	538.52
Other	2776.99	136	548.69
Total	2803.51	885	580.81

We want to test the null hypothesis that the average birth weight for Malay children is the same as that for Indian children. The test statistic is computed as below. The pooled standard deviation is:

$$s = \sqrt{[(457*597.52^2) + (134*538.52^2)]/134 + 457)}$$

$$= 584.66$$

$$t = (2816.71-2914.26)/\{584.66[(1/458)+(1/135)]\}$$

$$= -17.40$$

This should have a t-distribution with 591 degrees of freedom, if the null hypothesis is true. The t-distribution is approximately the same as the normal distribution when the sample size is large (more than 50). Thus, the cut point for a 5% level of significance would be ± 1.96, and since the calculated t is outside these limits, we would reject the null hypothesis and conclude that the two groups have different average birth weights.

[Note: Comparison of more than two groups would require more advanced statistical tests such as the Analysis of Variance and F-tests, which are beyond the scope of this manual. Refer to other statistical texts, such as Glantz, for details.]

Comparison of two means (paired samples)

As in the case of the McNemar's test, when the two samples are not independent (usually paired due to matching), a similar t-test can be computed. The procedure involves the computation of differences in the outcome variable between the two members of the pair, and calculating the mean and the standard error of these differences. The ratio, t = (mean difference / standard error of difference), then follows a t-distribution with (n-1) d.f., where n is the number of pairs, when the null hypothesis is true.

8.5 References and further reading

Fleiss J.L. Statistical methods for rates and proportions. New York, John Wiley and Sons, 1981.

Glantz S.A. Primer of biostatistics, 4 ed. Singapore, McGraw Hill, 1997.

Kahn H.A., Sempos C.T. Statistical methods in epidemiology. Oxford, Oxford University Press, 1989.

Kelsey J.L., Thompson W.D., Evans A.S. Methods in observational epidemiology. Oxford, Oxford University Press, 1986.

Kleinbaum D.G., Kupper L.L., Morganstern H. Epidemiologic research: principles and quantitative methods. New York, Van Nostrand Reinhold, 1982.

Lillienfeld A.M., Lillienfeld D.E. Foundations of epidemiology, 2 ed. Oxford, Oxford University Press, 1980.

Schlesselman J.J. Case control studies. Oxford, Oxford University Press, 1982.

Chapter 9

Association and Causation

9.1 Introduction

The most outstanding contribution of epidemiology is the study of association and causation in health and disease. Ironically, this is also the most difficult field in epidemiology, since it is often not easy to tell whether an observed association between a condition and a risk factor represents a cause-and-effect relationship.

The reasons for interest in establishing or excluding causality are:

- to understand the determinants of disease occurrence, distribution and outcome;

- to identify the links in the chain of causality that are amenable to intervention through general or specific intervention programmes; and

- to relate the output and impact of intervention programmes to their input, i.e. a causal evaluation.

9.2 Defining an association

An association is said to exist between two variables when a change in one variable parallels or coincides with a change in another. This is also called 'covariation' or 'correlation'. An association or covariation may be positive or negative and may be proportionate or disproportionate. An association is said to be <u>causal</u> when it can be proved that a change in the independent variable (exposure) produces (induces, results in, leads to, determines or causes) a change in the

dependent variable (disease). More appropriately, a causal relationship exists when exposure enters into the causation of disease. This underlines the possibility of multiple causes.

9.3 Defining the variables in an association

9.3.1. Independent and dependent variables

The hypothesis to be tested in a study usually defines which variable is assumed to be causal (i.e. is a risk factor) and which variable is considered to be the effect. The definition of a variable therefore depends on the study hypothesis: a variable may be independent in one hypothesis, a confounder in another, and dependent in a third. Take for instance, 'hypertension' in the simplified models shown in Figure 9.1.

FIGURE 9.1   VARIABLES INVOLVED IN HYPERTENSION
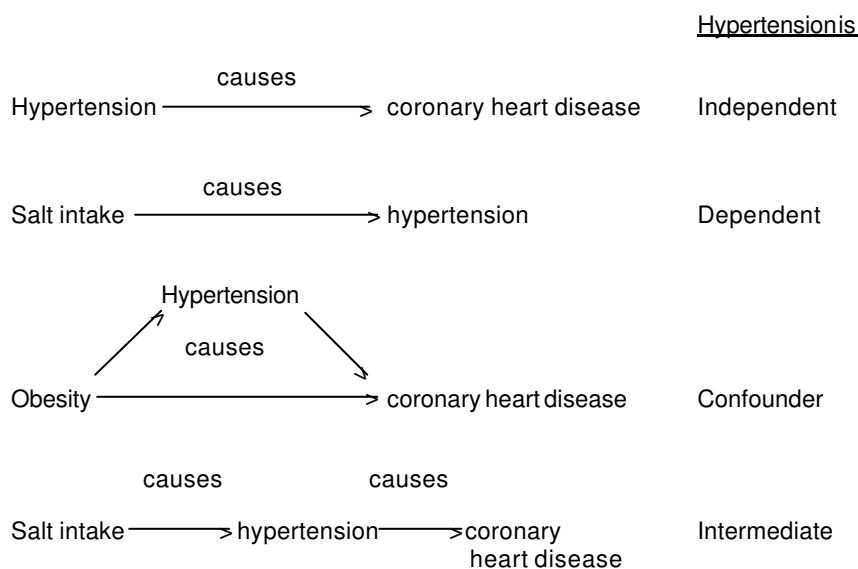
9.3.2. Confounding variables

A confounding variable is an independent variable (other than the hypothesized causal variable) that has or can have an effect on the dependent variable, but the distribution of which is systematically correlated with that of the hypothesized causal variable.

9.3.3. Control variables

Control variables are independent variables (other than the causal variable) which are potential confounders, and hence should be controlled or neutralized in the design or analysis. Note that these are only the 'known' or controllable variables. In most studies, it is impossible to control for all variables other than the suspected causal variables. By the process of randomization, it is hoped that many of these 'uncontrolled' variables will be equally distributed between the exposure and control groups.

9.3.4. Intermediate or intervening variables

When the effect of a causal variable on the dependent variable or study condition is mediated through a third set of variables, the latter are called intermediate variables. They are in fact dependent variables in relation to the causal variable, and independent in relation to the subsequent condition. Recall the example in Figure 9.1, in which the effect of salt on coronary heart disease was mediated through hypertension. The role of intermediate variables must be given more emphasis in epidemiology, especially when the hypothesized causal variables are global, such as, 'social condition' or 'development', when their association with, e.g. infant mortality is considered. In answering the question, 'How, in fact, does an increase in income or education bring about a reduction in infant mortality?' we should, for example, consider increased access to prenatal care, better maternal and infantile nutrition, access to vaccination, and better housing and personal hygiene. These are intermediate variables, some of which should be specified in the study design, and about which data are collected.

9.3.5. Effect modification

Some independent variables may modify (positively or negatively) the effect of the hypothesized causal variables. For example, hypertension is more frequent among black than among white Americans, while coronary heart disease is more frequent in whites

than in blacks. It is possible, therefore, that something related to the constitution or way of life of blacks modifies the effect of hypertension on coronary heart disease among them. Thus,

```
                              ⌐— black?
                            ⌐        ↓
Hypertension  ——————————————————————→ coronary heart disease
```
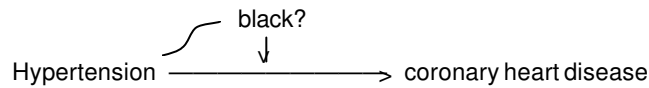
Some confounding variables are also effect modifiers.

9.4 Measuring an association

When the incidence (or prevalence) of a condition (e.g. lung cancer) in a group with certain characteristic (e.g. smoking) differs from the incidence (or prevalence) in a group without the characteristic (e.g. non-smokers), an association is inferred that may or may not be causal. The strength of the association is commonly measured by the relative risk or odds ratio (OR), in addition to attributable risk and population attributable risk per cent.

Another measure of association is the correlation between two variables. This can be expressed graphically in a correlation or scatter diagram (Figure 9.2), when the dependent variable (e.g. lung cancer incidence or mortality) is plotted on the vertical or Y-axis and the independent variable or characteristic (e.g. number of cigarettes smoked) on the horizontal or X-axis. If an association exists, changes in Y will coincide with changes in X. The relationship can also be expressed in terms of a correlation coefficient, or r, which is a measure of the degree to which a dependent variable varies with an independent variable. The correlation coefficient varies between +1 and -1. Table 9.1 gives approximate degrees of association corresponding to levels of r, subject, of course, to statistical tests of significance.

**TABLE 9.1 APPROXIMATE DEGRESS OF ASSOCIATION
CORRESPONDING TO LEVEL OF <u>r</u>**

<u>r</u>	Degree of association
± 1.0	Perfect
± 0.7 to ± 1.0	Strong
± 0.4 to ± 0.7	Moderate
± 0.2 to ± 0.4	Weak
± 0.01 to ± 0.2	Negligible
0.0	No association

The common correlation coefficients in use include:

- the Pearson product-moment correlation coefficient,

- the Spearman rank-order correlation coefficient, and

- the Kendall tau correlation coefficient.

<u>Regression coefficients</u> can also be used in measuring association. They are a measure of the mean changes to be expected in the dependent variable for a unit change in the value of the independent variable.

When more than one independent variable is associated with the dependent variable, multiple regression analysis will indicate how much of the variation observed in the dependent variable can be accounted for by one, or a combination of independent variables.

FIGURE 9.2   SCATTER DIAGRAMS AND CORRELATION COEFFICIENTS (r)
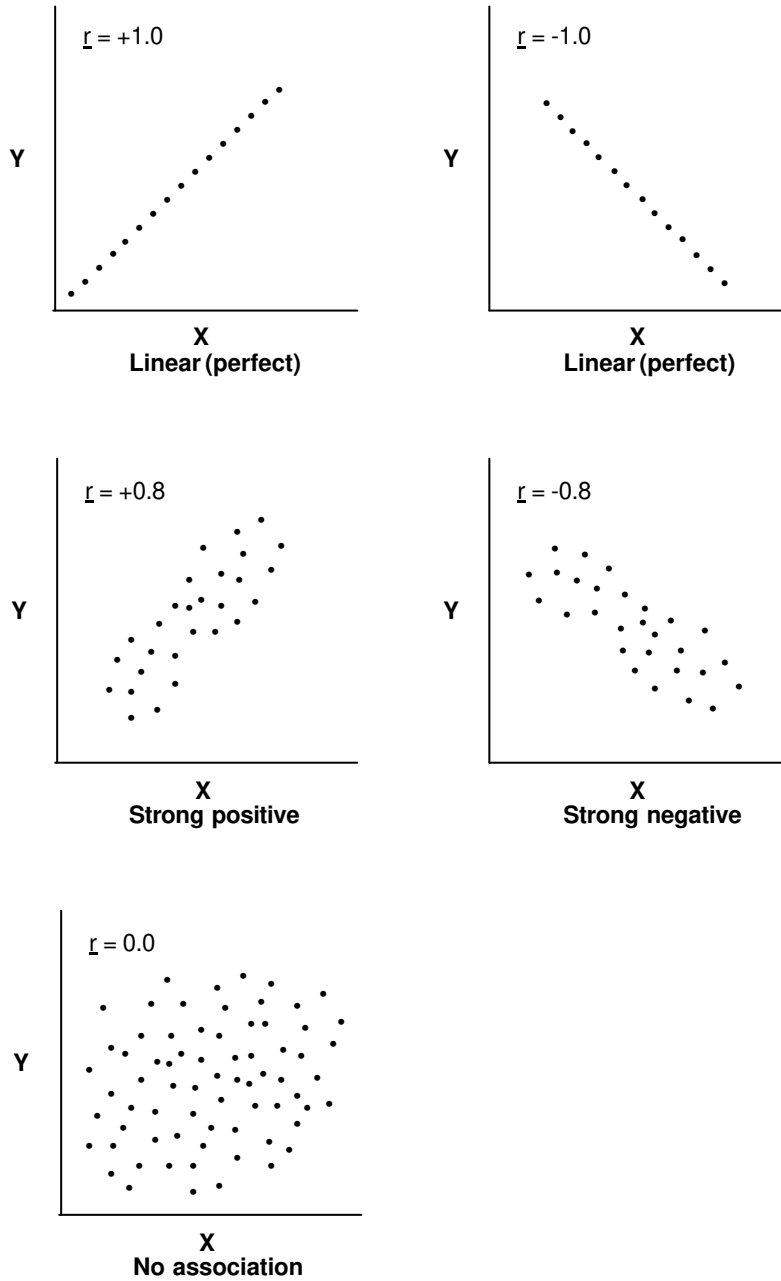
9.5 Problems in establishing causality

1. The existence of a correlation or association does not necessarily imply causation.

2. The concept of a single cause (the agent), once held in relation to communicable disease, has been replaced by the concept of multiple causation in diseases such as cancer and heart disease. Even in communicable diseases, factors in the agent, the host and the environment cooperate to cause the disease. For example, the tubercle bacillus is a necessary, but not a sufficient factor in the causation of tuberculosis.

3. The criteria used in establishing causality in infectious disease, namely, Koch's postulates, are not applicable to non-infectious diseases. Koch's postulates are:

    a. The organism is always found with the disease in accord with the lesions and clinical stage.

    b. The organism is not found in any other disease.

    c. The organism is isolated from one who has the disease, and cultured through several generations.

    d. The organism from culture is capable of producing disease in susceptible animals.

    Even in some infectious diseases, these postulates are not totally applicable.

4. The period between exposure to a factor or cause, and the appearance of clinical disease, is relatively long in non-infectious diseases. During this latent period, exposure to other factors complicates the research.

5. Specificity, easily established in infectious disease, does not apply to most other diseases. Lung cancer, for example, can result from smoking or exposure to radiation, asbestos or nickel dust. At the same time, each of these risk factors can cause diseases other than lung cancer. Smoking, for example, is involved in the causation of heart disease and emphysema.

6. Certain 'noise' factors, or confounders that are associated with the cause of a disease tend to distort or confound the relationship with the suspected factors. These require special handling during design or analysis to control or neutralize their effect.

7. Several systematic errors or bias in research design or data collection can produce false or spurious associations.

8. No statistical method can differentiate between causal and non-causal associations.

Because of these many uncertainties, the terms 'causal inference', 'causal possibility' or 'likelihood' are preferred to 'causal conclusion'. Such inferences would be enough in many situations to formulate policy rather than waiting for the unequivocal proof, which may be unattainable in several disease conditions.

9.6 Steps in establishing causality

Epidemiological strategies are usually assessed according to their power to provide a basis for causal inferences. It should be emphasized, however, that causal inference should not be made until certain requirements have been satisfied, which relate to two major questions:

• Is there actually an association?

• If there is an association, is it likely to be causal?

The requirements for making a causal interference aim (i) to exclude a non-causal association, and (ii) to ascertain the likelihood of a causal association. The requirements are given below:

1. The association actually exists and is statistically meaningful.

   This requires that:

   a. The association is not due to chance, as asserted by statistical tests of significance that can be applied to the difference between the frequency of the disease (the dependent variable) among those with and those without exposure to the risk factor (the independent variable). Tests can also be applied to the relative risk of disease in the two groups or to the correlation coefficient. Such tests would determine how frequently an association of the observed magnitude would occur solely on the basis of random variation or chance.

b. The association exists at the individual level and is not based only on the association measured on an ecological level, i.e. when the aggregate or geographical unit was used as the unit of observation. The possibility of ecological fallacy precludes inferring causality on an individual level.

c. The association is not based on numerator analysis, i.e. per cent distribution of 'cases' (the dependent variable), but on the appropriate population-based rates, calculating the relative risk or odds ratio.

2. The association is not spurious (i.e. not due to bias).

Spurious association can be of three types:

• due to selection bias,

• due to information or measurement bias, and

• due to confounding bias.

This will be discussed in detail below.

3. The confirmatory criteria for causality are satisfied.

Even if a statistical association does not exist and is not due to bias, a causal inference cannot be made confidently without satisfying the confirmatory criteria of causality. These relate to specific qualities of the association between the risk factor and the disease, namely, its strength, biological gradient, temporality, coherence, biological plausibility, specificity, consistency and experimental proof. These criteria are elaborated upon below.

The steps for establishing causality are represented diagrammatically in Figure 9.3.

Note: When intervening variables or mechanisms are involved, information on these variables should be collected as well.

9.7 Confirmatory criteria for a causal inference

Having established a statistical association and having ruled out sources of bias (i.e. having established that the association is not spurious), other specific criteria should be satisfied to support the causal inference.

FIGURE 9.3 - ESTABLISHING A CAUSAL INFERENCE

```
┌─────────────────────┐              ┌─────────────────────┐
│ Statistical          │    No        │ Non-causal           │
│ association          │─────────────▶│ association          │
│ established          │              │ (or repeat study     │
└─────────────────────┘              │ on a large sample)   │
         │                            └─────────────────────┘
       Yes
         │
         ▼
┌─────────────────────┐              ┌─────────────────────┐
│ Selection and        │    No        │ Non-causal           │
│ information          │─────────────▶│ association          │
│ bias excluded        │              │                      │
└─────────────────────┘              └─────────────────────┘
         │
       Yes
         │
         ▼
┌─────────────────────┐              ┌─────────────────────┐
│ Confounding excluded │    No        │ Non-causal           │
│ or neutralized and   │─────────────▶│ association          │
│ association persists │              │                      │
└─────────────────────┘              └─────────────────────┘
         │
       Yes
         │
         ▼
┌─────────────────────┐              ┌─────────────────────┐
│ Confirmatory criteria│    No        │ Non-causal           │
│ of causality (strength,│───────────▶│ association          │
│ consistency, specificity,│          │                      │
│ temporality, plausibility,│         └─────────────────────┘
│ experimental proof)  │
│ satisfied            │
└─────────────────────┘
         │
       Yes
         │
         ▼
┌─────────────────────┐   Specify    ┌─────────────────────┐
│ CAUSAL INFERENCE     │─────────────▶│ Causal model         │
│                      │              │ Direct               │
└─────────────────────┘              │ Indirect             │
                                      │ Interaction          │
                                      └─────────────────────┘
```

The association is strong (strength).

The strength of the association is measured by the relative risk (and attributable risk) and OR (in case-control studies). Correlation and regression coefficients can endorse these measures of effect. The stronger the association, the higher the likelihood of a causal relationship.

There is biological gradient.

A dose-response relationship (if present) can increase the likelihood of a causal association. This is not, however, possible in all studies.

The association follows a time sequence (temporality).

It goes without saying that the risk factor or cause must precede the condition or effect. This antecedent-consequence requirement is often overlooked. It is easier to establish temporality in experimental and cohort studies than in case-control and cross-sectional studies.

The association is plausible (coherence or plausibility).

The association should make common biological or sociological sense and should not conflict with existing theories or knowledge unless it is actually a challenge to those theories. In either case, there should be some theoretical basis explaining the association.

The association is consistent (consistency).

Causality is more likely when the association is supported by other investigations conducted by different persons in different places, circumstances and time-frames, and using different research designs.

The association is specific (specificity).

The disease outcome should be specific to, or characteristic of, exposure to a particular risk factor. This is more feasible in infectious diseases than in non-infectious diseases, which can result from different risk agents. Hence, this criterion is not generalized.

There is experimental proof for causality.

Two types of experimental proof can be established: (i) experiments in humans using the risk factor, which are difficult to provide, and (ii) cessation experiments, whereby removal of the putative cause results in a significant reduction in disease incidence.

9.8 Types of association

The association between two variables may be causal or non-causal.

9.8.1. Causal association

As already stated, a causal association exists when the independent variable (risk factor) causes changes in the dependent variable. Causal associations are of three types (see Figure 9.4).

a. Direct causal association

    A direct causal association is inferred when the risk factor or independent variable changes the dependent variable or condition directly, without intervening variables, e.g. exposure to the tubercle bacillus causes tuberculosis, exposure to lead causes lead poisoning, and iodine deficiency causes goitre.

b. Indirect causal association

    The association is inferred when the risk factor or independent variable causes changes in the dependent variable or condition through the mediation of other intermediate variables or conditions:

    iodine deficiency ———————→ goitre ———————→ thyroid adenoma

    Thus, thyroid adenoma is caused indirectly by iodine deficiency. Note that the term 'indirect association' may be used in a broader sense. For example, endemic goitre is associated with high altitude simply because water supplies are likely to contain less iodine at high rather than low altitudes. Such usage, however, should be restricted and carefully evaluated. The main issue here is whether the association is causal or non-causal. The criteria for causality should apply equally to direct and indirect causal associations.

FIGURE 9.4   CAUSAL MODELS

If E is the exposure factor and D is the disease   <u>Causality</u>

1.   E —————————→ D   Directly causal

2.   E —→ C —————————→ D   Indirectly causal

3. E_1 ⟍
                    ⟩ D   Independently causal
      E_2 ⟋   (if rate increases,
                                                    there is synergism)

4. E_1 ⌐   Conditionally causal
            |————→ D   (only when $E_1 + E_2$
      E_2 ⌐   are present)

          E_2
5. E_1 —————————→ D   Effect modification
                                                    (or form of synergism)

           E_2
6. E_1 —————————→ D   Confounding association of E_1
                                                    and D disappears by
                                                    neutralization or stratification
                                                    analysis

7.   ⟋→ D_1 Leukaemia

      E ⟨———————————→ D_2 Lung cancer   One cause with
      Radiation   multiple effects
                      ⟍→ D_3 Radiation
                                  sickness

8. E_1 Radiation ⟍
      E_2 Nickel   ⟍   One disease caused
                         ⟩ D   Lung cancer   by multiple factors
      E_3 Smoking   ⟋   (independently), as #3
      E_4 Asbestos ⟋

9.   ⟋→ D_1 Heart disease   Hidden third variable:
                                                     heart disease and
      Stress   gastric ulcer may occur
      E_3 ⟨   concurrently (i.e. are
                                                     associated) because both are
                      ⟍→ D_2 Gastric ulcer   related to stress

c. Interaction (including conditional) causal association

There may be interactions (positive or negative) between categories of independent variables that produce changes in the dependent variables.

One form is synergism (or antagonism) between two variables, whereby each factor has an independent effect on the condition, while the joint effect is greater (or smaller) than each alone. In one-way analysis, each factor has an effect on the condition:

$$X_1 \longrightarrow Y$$

$$X_2 \longrightarrow Y$$
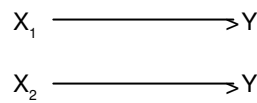
In stratification analyses (e.g. control table analysis), neither effect disappears, but the joint effect may be greater (or smaller):
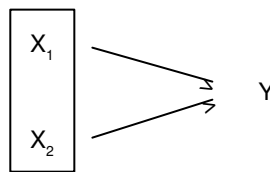
For example, measles can result in death, but the probability is greater in malnourished children:
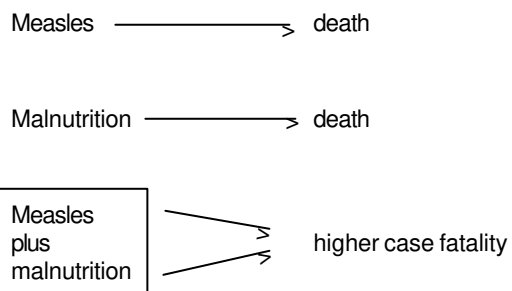
In a <u>conditional causal association</u>, two risk factors are incapable of producing a condition unless they exist in the presence of each other. For example, blackwater fever (a febrile condition characterized

by dark urine due to haemolysis) follows malaria as a complication only if the malaria was due to *Plasmodium falciparum* and the cases were treated with quinine.

Falciparum malaria alone ⎯⎯⎯⎯⎯⟶ no blackwater fever

Quinine alone ⎯⎯⎯⎯⎯⟶ no blackwater fever

Falciparum malaria

plus ⎯⎯⎯⟶ blackwater fever

quinine use
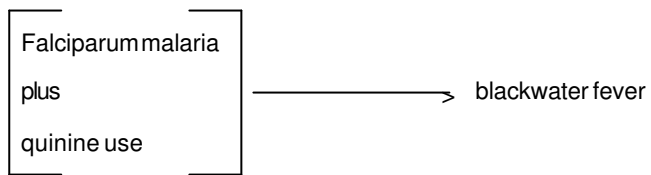
9.8.2. Non-causal, spurious association

In some situations, an association does exist, but, despite its significance and strength, it may be spurious or non-causal as far as the special characteristics under study are concerned. A non-causal association is inferred when the association is:

- due to chance,

- based on numerator analysis or ecological correlation, or

- due to bias.

9.9 References and further reading

Hill A.B. Principles of medical statistics. New York, Oxford University Press, 1966.

Kelsey L.T., Thompson W.D., Evans S.A. Methods in observational epidemiology. New York, Oxford University Press, 1986.

Kleinbaum D.G., Kupper L.L., Morganstern H. Epidemiologic research: principles and quantitative methods. New York, Van Nostrand Reinhold, 1982.

Lillienfeld A.M., Lillienfeld D.E. Foundations of epidemiology, 2 ed. New York, Oxford University Press, 1980.

Chapter 10

Ethical Aspects of Health Research

10.1 Introduction

The application of experimental methods to biomedical research is a product of the 20th century. Many fundamental discoveries were made before this time, but progress was subsequently achieved through the application of scientific principles to medical and public health practices.

During almost the whole of human history, the only drugs used were naturally-occurring substances of animal, vegetable or mineral origin, and long experience had shown that, in the doses used, they did no serious harm (and, in most cases, not much good either). However, a century ago, the chemical industry started to develop, for medical use, synthetic compounds that had never existed in nature. The first of these to have an important impact on the treatment of human disease was Salvarsan (arsphenamine), introduced primarily as a remedy for syphilis.

An experiment is an attempt to discover something unknown, or to test a supposition or principle, but we cannot be sure of the outcome. By definition, an experiment involves chance. It is because of this chance or element of the unknown that ethics become a paramount issue in those experiments which involve human subjects. Much basic and developmental biomedical research could be undertaken successfully on animal models; however, absolute reliance cannot at present be vested in these models as indicators of physiological, pharmacological or toxicological response in man. All innovative scientific interventions, whether diagnostic, prophylactic

or therapeutic, should ultimately be evaluated in human subjects. The need for safeguards in human experimentation cannot be overemphasized, and several important codes have been developed for the protection of human subjects.

The three underlying principles are:

1. beneficence, which requires that good should result, harm should be avoided, or that benefits should justify the expected risk or harm;

2. respect for rights, including the free choice of the subject and protection for those of diminished autonomy; and

3. justice, which requires an equal distribution of burden and benefit.

10.2 International declarations

The first important code of ethics was the Nurenberg Code of 1947: no research could proceed on human subjects without 'voluntary consent', and this has remained unchanged in subsequent codes.

The World Medical Association, assisted by WHO, developed an expanded and revised code of ethics to guide doctors in research involving human subjects, called the Declaration of Helsinki. This was followed by a revised Declaration in 1975 (Helsinki II), which changed the emphasis from 'clinical research' to 'biomedical research involving human subjects'. This was adopted at the 29th World Medical Assembly in Tokyo in 1975.

The demands for new and better treatment, and its greater distribution, have vastly multiplied the demands for biomedical research involving human subjects – especially clinical trials. In the regulation of trials and other biomedical research involving human subjects, processes of review have been developed by governmental and institutional boards and committees, which draw heavily upon the guidelines of the Helsinki codes, including, particularly, the following:

*   Biomedical research should follow scientific principles and should be based on adequately performed laboratory and animal experimentation and on a thorough knowledge of the scientific literature.

*   The design of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol, to be reviewed by an independent committee.

- The experiment should be conducted by scientifically qualified person(s) and under the supervision of clinically competent medical experts.

- Biomedical research involving human subjects cannot legitimately be carried out unless the importance of the objectives can justify the inherent risk to the subject.

- Every biomedical research project involving human subjects should be preceded by careful assessment of predictable risks in comparison with foreseeable benefits to the subject or others. Concern for the interests of the subject must always prevail over the interests of science and society.

- The right of the research subject to safeguard his or her integrity must always be respected. Every precaution should be taken to respect the privacy of the subject and to minimize the impact of the study on the subject's physical and mental integrity and on his or her personality.

- The accuracy of research results must be preserved.

- In any research on human beings, each potential subject must be adequately informed of the aim, methods, anticipated benefits and potential hazards of the study and the discomfort it may entail.

- When obtaining informed consent for a research project, a doctor should be particularly cautious if the subject is in a dependent relationship to him or her. No pressure or threat should be exercised.

- In case of legal incompetence, informed consent should be obtained from the legal guardian in accordance with national legislation.

- Subjects should be informed that they are free to abstain or to withdraw from participation at any time.

Of itself, however, informed consent offers an imperfect safeguard to the subject, and it should always be complemented by independent ethical review of research proposals. Moreover, many individuals, including children and adults who are mentally ill or defective, or who are totally unfamiliar with modern medical concepts, are therefore incapable of giving adequate consent. For such groups, in particular, independent ethical review is imperative.

10.3 Consent of subjects

1. Children

   It is axiomatic that children should never be the subjects of research that might equally be well carried out on adults. However, their participation is indispensable for research on diseases of childhood and conditions to which children are particularly susceptible. The consent of a parent or other legal guardian, after a full explanation of the aims of the experiment and of possible hazards, discomfort or inconvenience, is always necessary.

2. Pregnant and nursing women

   Pregnant and nursing mothers should, under no circumstance, be the subjects of non-therapeutic research that carries any possibility of risk to the fetus or neonate, unless this is intended to elucidate problems of pregnancy or lactation. Therapeutic research is permissible only with a view to improving the health of the mother without prejudice to that of the fetus, to enhancing its viability or to aiding the nursling's healthy development, or the ability of the mother to nourish it adequately.

   Research directed to induced termination of pregnancy, or undertaken in anticipation of termination, is an issue that is dependent upon national legislation and religious and cultural precepts, and therefore does not lend itself to an international recommendation.

3. Mentally ill and mentally defective persons

   Substantially similar ethical considerations apply to the mentally ill and the mentally defective. They should never be the subjects of research that might equally well be carried out on adults who are in full possession of their mental faculties. They are, however, the only subjects available for research on the origin and treatment of mental disease or disability.

   The agreement of the immediate family -- whether spouse, parent, adult offspring or sibling -- should be sought, but it is sometimes of doubtful value, especially as mentally deranged or defective patients are sometimes regarded by their families as an unwelcome burden.

4. Other vulnerable social groups

The quality of the consent of subjects who are junior or subordinate members of a hierarchically structured group requires careful consideration, as willingness to volunteer may be unduly influenced by the expectation of adventitious benefits. Examples of such groups are medical and nursing students, subordinate laboratory and hospital personnel, employees of the pharmaceutical industry, and members of the armed forces. More seriously objectionable is experimentation on exclusively selected national or cultural groups.

5. Community-based research

When research is undertaken on a community basis --for example by experimental treatment of water supplies, health systems research, large-scale trials of new insecticides, and nutritional fortification or substitutes -- individual consent on a person-to-person basis may not be feasible, and the ultimate decision to undertake the research rests with the responsible public health authority.

Nevertheless, all possible means should be used to inform the community concerned of the aims of the research, the advantages expected from it, and any possible hazards or inconveniences.

10.4 Review procedures

In highly centralized administration, a national review committee may be constituted to review research protocols from both scientific and ethical standpoints. In countries where medical research is not centrally directed, protocols are more effectively and conveniently reviewed from the ethical standpoint at local or regional level. The basic responsibilities of locally operative ethical review committees are twofold:

- to verify that all proposed interventions and, particularly, the administration of drugs under development have been assessed by a competent expert body as acceptably safe to be undertaken in human subjects; and

- to ensure that all other ethical considerations arising from a protocol are satisfactorily resolved, both in principle and in practice.

Whatever the pattern of the procedure adopted for ethical review, it should be based on a detailed protocol comprising the steps outlined in this manual. Care should be taken to ascertain the criteria for determining admission and withdrawal of individual subjects, including full details of the informed consent procedure.

Information should also be included to establish:

- the safety of each proposed intervention and of any drug or device to be tested, including the results of relevant laboratory and animal research;

- the presumed benefits and potential risk of participation;

- the means proposed to elicit informed consent, or, when this is not possible, satisfactory assurance that the guardian or family will be appropriately consulted and the rights and welfare of each subject will be adequately protected;

- that the investigator is appropriately qualified and experienced, and commands adequate facilities for the safe and efficient conduct of the research;

- that provisions will be made to protect the confidentiality of the data; and

- the nature of any other ethical considerations involved, together with an indication that the principles enunciated in the Declaration of Helsinki will be implemented.

10.5 References and further reading

Bankowski Z., Bernardelli J.C. Medical ethics and medical education. Geneva, Council for International Organizations of Medical Sciences, 1981 (CIOMS round table proceedings, no. 14).

Bankowski Z., Howard-Jones N. Biomedical research involving animals. Geneva, Council for International Organizations of Medical Sciences, 1984 (CIOMS round table proceedings, no. 17).

Proposed international guidelines for biomedical research involving human subjects. Geneva, Council for International Organizations of Medical Sciences, 1982.

Chapter 11

Construction of a Research Proposal

11.1 Stepwise construction of a research design using the WHO/WPRO format for a research proposal

A copy of the WHO/WPRO application form for financial support for health research is attached as an Appendix.

11.2 Statement of problem

The first step in the development of a research project is to state the research problem in precise, clear terms. Definition of the research problem is a prerequisite to clarifying and focusing the investigator's chosen topic.

The statement of the problem:

- is the essential basis for the construction of a research proposal (research objectives and hypotheses, methodology, work plan and budget, etc.);

- is an integral part of selecting a research topic;

- will guide and put into sharper focus the research design being considered for solving the problem;

- allows the investigator to describe the problem systematically, to reflect on its importance, its priority in the country and in the local area, and to point out why the proposed research on the problem should be undertaken;

- facilitates a peer review of the research proposal by funding agencies.

How should the statement of the problem be written for a research proposal? The writing should be precise and concise but should include essential points. Information about the problem should be summarized so that the reader is not 'drowned' in detail.

11.3 Relevance of the problem to national or local health activities (biomedical, behavioural and health systems development)

Why the proposed research on the problem should be undertaken, and the importance of the problem and its priority in the country and in the local area, should be spelled out. A description of the problem is important, with a statement about how the results will be used.

11.4 Field(s) of application of the proposed results

Describe how the results of the study will be useful for policy-makers, health administrators or health scientists, and how the results will be transmitted to the appropriate audience.

11.5 Review of literature and other existing information

The second step is for investigators to familiarize themselves with existing knowledge about the research problem and to find out whether or not others have investigated the same or similar problems. This step is accomplished by a thorough and critical review of the literature and by personal communication with experts. A review of existing information is important when preparing a proposal because:

- it helps further understanding of the problem proposed for research, and may lead to refining of the 'statement of the problem';

- it helps to identify the study variables and conceptualize their relationships;

- it helps in the formulation and selection of research hypotheses;

- it helps in finding out what others have reported on the topic, so that account can be taken of this design of the research;

- it provides familiarity with the various methods that might be used in the research.

The source of information may include the following:

- card catalogues of books in libraries;

- indices, such as the <u>Index Medicus</u> and the <u>International Nursing Index</u>, which identify journal articles by subject, author and title;

- computer-based literature searches such as MEDLINE, MEDLARS and CATLINES;

- bibliographies, such as those found at the end of books, articles and theses, or prepared as separate documents;

- statistics collected at national, provincial and/or departmental levels; and

- responses to enquiries about ongoing research.

11.6 Statement of objectives

Research objectives are the goal to be achieved by a research project. Differentiation between 'general' and 'specific' objectives may eliminate unnecessary confusion. The <u>general objective</u> of research is what is to be accomplished by the research project and why.

> <u>Example</u>: to determine whether or not a new vaccine should be incorporated into public health programmes.

The <u>specific objectives</u> are, in detail, the specific aims of the research project, often breaking down what is to be accomplished into smaller logical components. In other words, specific objectives relate to the specific research questions the investigator wants to answer through a proposed study.

> <u>Example:</u> in evaluating a new vaccine, to determine the degree of protection that is attributable to the vaccine in a study population by comparing the vaccinated and unvaccinated groups.

11.7 Variables

It is necessary to identify the variables that will be involved in the research project being designed. Four types of variable are important in research:

a. <u>Independent variables:</u> variables that are manipulated or treated in a study in order to see what effect differences in them will have on those variables proposed as being dependent on them.

   <u>Synonyms:</u> cause, input, predisposing factor, antecedent, risk factor, characteristic, attribute, determinant

b. <u>Dependent variables:</u> variables in which changes are results of the level or amount of the independent variable or variables.

   <u>Synonyms:</u> effect, outcome, consequence, result, condition, disease

c. <u>Confounding or intervening variables:</u> variables that should be studied because they may influence or 'confound' the effect of the independent variable(s) on the dependent variable(s). For instance, in a study of the effect of measles (independent variable) on child mortality (dependent variable), the nutritional status of the child may play an intervening role.

d. <u>Background variables:</u> variables that are so often of relevance in investigations of groups or populations that they should be considered for possible inclusion in the study.

   <u>Synonyms:</u> sex, age, ethnic origin, education, marital status, social status

The objective of research is usually to determine the effect of changes in one or more independent variables on one or more dependent variables. For example, a study may ask 'Will alcohol intake (independent variable) have an effect on development of gastric ulcer (dependent variable)?'

Certain variables may not be easy to identify. The characteristics that define these variables must be clearly identified for the purpose of the study. During the planning stage, the variables in a study should be clearly identified and their method of measurement, as well as the unit of measurement, clearly indicated.

11.8 Statement of research hypothesis

The value of scientific work depends heavily on the originality and logic with which hypotheses are formulated. If researchers know enough to make predictions concerning what they are studying, hypotheses may be formulated. A hypothesis can be defined as a tentativeprediction or explanation of the relationship between two or more variables. A hypothesis, in other words, translates the problem statement into a precise, unambiguous prediction of expected outcomes. It must be emphasized that hypotheses are not meant to be haphazard guesses, but should reflect the depth of knowledge, imagination and experience of the investigator. A hypothesis can be as simple in form as predicting the relationship between two variables, one <u>independent</u> and one <u>dependent</u>. Therefore, in the process of formulating hypotheses, all variables relevant to the study should be identified.

<u>Example</u>: Health education involving active participation by mothers will produce more positive changes in child feeding than health education based on lectures.

Independent variable: types of health education

Dependent variable: changes in child feeding

11.9 Research methodology

a. Summary of methodology (not more than 150 words)

Give one or two paragraphs summarizing the salient points of the research design.

b. Research design

(1) Selection of research strategies

The selection of a research strategy is the core of research design, and is probably the single most important decision the investigator has to make. The choice of strategy, whether descriptive, analytical, experimental, operational, or a combination of these, depends on a number of considerations. The specific types of studies are as follows:

<u>Descriptive strategies</u> (observational hypothesis generation rather than testing)

- descriptive cross-sectional study or population survey, e.g. malaria survey, opinion survey, knowledge-attitude-practice (KAP) survey;

- epidemiological description of disease occurrence by person, place and time;

- studies of changing patterns of health and disease over time and space: the epidemiological translation;

- community diagnosis of a health problem or assessment of needs;

- studies of existing data: case-series, disease registries, surveillance reports;

- studies of the natural history of disease.

<u>Observational analytical strategies</u> (hypothesis testing)

- prospective study (cohort study);

- historical (or reconstructed) cohort study, when adequate historical data or records are available;

- retrospective study (case-control study);

- analytical cross-sectional study;

- follow-up study (longitudinal study; repeated cross-sectional study).

<u>Experimental strategies</u>

- animal studies;

- therapeutic clinical trials;

- prophylactic clinical trials;

- field trials;

- quasi-experimental studies (intervention studies, health systems research).

<u>Operational strategies</u> (observation, time-motion study)

(2)   Selection of research setting

The research setting includes all the pertinent facets of the study, such as the population to be studied, the place and time of the study, and consideration of ethical problems.

(3)   Sampling

Sampling is the process or technique of selecting a sample of appropriate and manageable size for study. In epidemiological investigations, it is almost always possible to deal with a sample drawn from a reference population or universe. The universe may be a population of people (healthy and sick), a population of cases of a certain disease, or recipients of a certain treatment.

- Selection of probability sampling method: simple random; systematic and stratified sampling; cluster sampling; multiphasic; multistage; sequential; repetitive; weighted and stratified.

- Determination of sample size: the sample should be of sufficient size to produce meaningful results and to allow tests of statistical significance to be applied.

- Plans should be made to ensure representativeness and reliability of the sample to minimize sampling errors.

(4)   Use of controls

Control or comparison groups are used in scientific research in order to increase the validity of the conclusions. Control groups consists of comparable units from the same population, but who differ in some respects, namely in exposure to risk factors, use of a preventive or therapeutic measure, or participation in an intervention programme.

In an experimental study, the control group consists of those subjects to whom no experimental stimulus is administered, but who resemble members of an experimental group in all other respects. The subjects who form the experimental and control groups should be selected and allocated randomly to each group, if possible.

Control groups are not necessary in studies in which no attempt is made to show a cause-and-effect relationship, or to show that a certain result was due to a particular treatment or intervention. While some descriptive studies (studies of existing data, surveys) may lack control groups, they are necessary in all analytical epidemiological studies, in experimental studies of drug trials, in research on the effects of intervention programmes and disease control measures, and in many other investigations. Many gross errors have been made in attempting to equate groups and make generalizations based on comparisons between groups that, in reality, are very different. Therefore, plans must be made for testing equality between experimental (or sample) and control groups.

(5) Study instrument(s)

Instruments are tools by which data are collected. They include:

(a) questionnaire and interview schedules (see Annex 1):

- preparation, precoding and pretesting of questionnaires;

- plan for interviews and call-backs;

- preparation of instructional manual;

- training of interviewers.

(b) other methods of observation:

- medical examination;

- laboratory tests;

- screening procedures.

(c) design of recording forms.

(6) Short description of plans for collecting data:

(a) organization of study and data collection in order to minimize the possibility of confusion, delays and errors;

      (b)   organization and training of the data-collecting team and definition of responsibilities in the proposed study;

      (c)   logistic support for data collection;

      (d)   plans for test or feasibility studies, including pretesting methods; and

      (e)   plans for collaboration between different institutions, if applicable.

(7)   Short description of plans for analysis of data and interpretation of results

Plans for analysis are an integral part of the research design, and should be incorporated into the research proposal. Preparing such plans helps the investigator avoid several pitfalls, such as discovering at the end of the study that vital information has not been collected, that some of the information collected will not be included in the analysis, or that some of the information collected has not been gathered in a form appropriate for statistical analysis.

The description should include:

(a)   design of analysis form;

(b)   plans for processing and coding data, by manual sorting, machine sorting, computer programme or record linkage; and

(c)   choice of statistical methods to be applied to each hypothesis.

11.10 Example of a project description

Title of project

An epidemiological study of vasectomy and atherosclerotic diseases

1. Statement of problem

Vasectomy, which is a safe, simple and highly effective contraceptive method, has been widely performed throughout the world. In country A, more that one million, or about 8% of all males

of reproductive age underwent vasectomy during the period 1960-1985. Reports of studies in experimental animals in the USA in the late 1970s and early 1980s suggested that vasectomy may accelerate the progress of atherosclerosis. Understandably, these reports caused concern to vasectomy service providers, as well as to their past and prospective clients. The important question is whether the alleged association between vasectomy and atherosclerosis applies to human beings.

2. Relevance of the problem to national or local health objectives (biomedical, behavioural and health systems development)

In view of the worldwide publicity about the experimental findings in the lay press, and the negative impact they may have on vasectomy programmes, there are both programmatic and scientific reasons for conducting epidemiological studies on this problem. In selecting a developing country for the study, several factors had to be considered: the prevalence of vasectomies, the incidence of atherosclerotic disease, the number of years that vasectomy has been widely available, general access to medical services, and consistency of diagnostic skills. Analysis of relevant data showed that country A was most suitable on these grounds, and we decided to undertake the study in this country.

3. Field(s) of application of the proposed research results

Depending on the answers to the following questions, the study results will be useful in assisting family planning policy-makers and health scientists to implement vasectomy programmes in a more effective, safe way.

a. Is vasectomy associated with atherosclerotic diseases?

b. If an association exists, what is the relative importance of vasectomy in comparison with other, known risk factors?

c. What subgroups of men might be at special risk of developing atherosclerotic diseases following vasectomy?

4. Review of literature and other existing information

Twelve epidemiological studies have been conducted in the USA, the United Kingdom and northern European countries, none of

which has detected a causal association between vasectomy and cardiovascular morbidity and/or mortality in men. In most of these studies, however, the subjects had been vasectomized less than ten years before enrolment, while the latent period for cardiovascular disease may very well be considerably longer. Furthermore, different results might be obtained in other socioeconomic and cultural settings.

The major references are as follows:

Alexander, N. J. and Clarkson, T. B. (1978) Vasectomy increases the severity of diet-induced atherosclerosis in *Macaca fascicularis. Science, 201*, 538-541.

Clarkson, T. B. & Alexander, N. J. (1980) Long-term vasectomy effects on the occurrence of atherosclerosis in rhesus monkeys. *J. Clin. Invest., 65,* 15-25.

Goldacre, M. J., Clarke, J. A., Heasman, M. A. and Vessey, M. P. (1978) Follow-up of vasectomy using medical record linkage. *Am. J. Epidemiol., 108,* 176-180.

Walker, A. M., Jick, H., Hunter, J. R., Danford, A., Watkins, R. N., Alhadeff, L. and Rothman, K. J. (1981) Vasectomy and nonfatal myocardial infarction. *Lancet, ii,* 13-15.

Walker, A. M., Jick, H., Hunter, J. R., Danford, A. and Rothman, K. J. (1981) Hospitalization rates in vasectomized men. *J. Am. Med. Assoc., 245,* 2315-2317.

Wallace, R. B., Lee, J., Gerber, W. L., Clarke, W. R. and Lauer, R. M. (1981) Vasectomy and coronary disease in men less than fifty years old: Absence of an association. *J. Urol., 126,* 182-184.

5.   *Statement of objectives*

       a.   General objectives

             To determine whether there is a causal association between a vasectomy and subsequent hospitalization due to atherosclerotic diseases, and, if so, whether vasectomy potentiates the risk in subjects with other predisposing risk factors for coronary disease, such as smoking, hypertension and high cholesterol.

     b.   Specific objectives

- to estimate the overall relative risk of vasectomy, as well as other risk factors for atherosclerotic diseases in men (using a univariate method);

- to estimate the independent effect of vasectomy on atherosclerosis (using a conditional logistic regression model);

- to test the possible duration of the effect of vasectomy on risk for atherosclerosis; and

- to examine the possible synergistic effect between vasectomy, cigarette smoking and hypertension.

6. Variables

a.   Atherosclerotic diseases will be identified according to the WHO criteria.

b.   Patient characteristics: age, birth, date, religion, education, occupation, family history, marital status.

c.   Reproductive history: number and sex distribution of living children, wife's reproductive status.

d.   Lifestyle: smoking status, alcohol intake, dietary habits, salt intake, coffee drinking, physical activity.

e.   Medical history: diseases or operations that might have affected sterility, hypertension, diabetes or hypercholesterolaemia.

7. Statement of research hypotheses

Reports of studies on experimental animals in the USA in the late 1970s and early 1980s suggest that vasectomy may accelerate the progress of atherosclerosis. We wish to investigate whether this finding applies to human beings.

8. Research methodology

a.   Summary of methodology (not more than 150 words)

A hospital-based case-control study will be conducted to examine the possible relationship between vasectomy and atherosclerotic morbidity in men. Five hundred men aged 35-64 (cases) who were admitted to ten university-

affiliated hospitals and diagnosed for the first time with atherosclerotic disease will be compared with 1000 matched non-atherosclerotic patients (controls) hospitalized with a diagnosis considered to be unrelated to vasectomy.

b.   Research design

(1)   Selection of research strategies

The selected cases and controls as defined above will be interviewed by a trained interviewer using a pre-constructed questionnaire.

(2)   Selection of research setting

Recruitment of study subjects will be carried out in ten selected teaching hospitals in the country. The subjects must be currently married male patients, aged 35-64 years, with at least one living son. Their wives must not be sterile due to any medical condition during their reproductive period. Cases will be men hospitalized with a diagnosis, established before discharge, of a first episode of atherosclerotic disease. Controls will be male hospital admissions who have no history of atherosclerosis, and who were admitted with a disease not suspected of being related to vasectomy. Study period: September 1988-March 1990.

(3)   Sampling

In view of the design of this study, the sample (or cases and controls) will be selected by a non-randomized method. The sample size was set at 500 cases and 1000 controls by a sample size determination technique on the basis of a pre-assigned significance level and power, and the level of relative risk to be detected. An attempt will be made to avoid or diminish potential sources of bias and error that are frequently encountered in case-control studies. Misdiagnosis bias, recall bias, selection bias and vasectomy reporting bias will be of great importance to the validity of the study results.

(4)   Use of controls

Two controls (as previously defined) will be matched with each case by: (i) hospital (same); (ii) age (± 5 years); (iii) number of living children (at least one son); and (iv) admission date (closest). The diagnoses of controls will include: digestive system diseases, neoplasms, injury, poisoning, infectious or parasitic diseases, respiratory system diseases, nervous system and musculoskeletal diseases and others.

(5)   Study instrument(s)

The questionnaire will be structured to minimize interviewer and respondent bias. It will include questions on: (i) patient characteristics; (ii) family health history; (iii) reproductive history, (iv) habits; (v) personality type; (vi) medical history (including questions on vasectomy); and (vii) clinical information (from medical charts).

(6)   Short description of plans for collecting data

Recruitment of cases: interviewer reviews daily inpatient status on blackboard ⇒ if diagnosis falls into study category, refers case to chief cardiologist for review ⇒ doctor decides on eligibility of case ⇒ interviewer checks eligibility of patient's background ⇒ if the patient meets the eligibility criteria for diagnosis and background ⇒ interviewer performs interview and fills in questionnaire ⇒ on completion of a batch of five cases and ten matched controls, interviewer contacts research headquarters staff for review of questionnaires ⇒ repeats above procedure.

Recruitment of controls: interviewer reviews admission log and selects potential controls per case who fulfil matching criteria and have appropriate admission diagnoses ⇒ checks eligibility criteria of patient's background ⇒ if selected as eligible control, performs interview.

(7) Short description of plans for analysis of data and interpretation of results

Independent variables will dichotomized as follows: age, 35-54 <u>versus</u> 55-64; education, ≤ 12 <u>versus</u> ≥13 years; occupation, administrative <u>versus</u> others; cigarette smoking, ever <u>versus</u> never; coffee drinking, every day <u>versus</u> less often or never; history of physician-diagnosed diseases, present <u>versus</u> absent.

Data processing will be aided by computer.

Statistical analysis: odds ratios will be calculated for matched triplets (one case, two controls).

- Unadjusted odd ratios and their 95% confidence intervals will be calculated, using a univariate method allowing for matched sets, to estimate the overall relative risk of vasectomy as well as other risk factors.

- Independent effect of vasectomy on atherosclerosis will be evaluated by adjusted odds ratios, using a conditional logistic regression model allowing for matched sets.

- Effect of time elapsed since vasectomy on risk for atherosclerosis will be tested by a conditional logistic regression model, with atherosclerosis as the dependent variable, and the interval since vasectomy as the independent variable.

APPENDIX

WORLD HEALTH ORGANIZATION
Regional Office for the Western Pacific

APPLICATION FOR FINANCIAL SUPPORT FOR HEALTH RESEARCH

I. SUMMARY SHEET

1		Principal investigator Prof/Dr/Mr/Mrs/Miss
		Family name (surname)
		First name Other names
	1.1	Title of post, position or appointment presently held by principal investigator
	1.2	Complete postal address
2		Institution responsible for the research programme
		Name and address
3		Title of health research proposed (should be brief, precise and informative to workers outside your field)
	3.1	Objectives of research proposed (clearly and concisely list general and specific purposes of the proposed study)
	3.2	Duration of research
		From (date)
		To (date) Total (years)

	3.3	Funds required (US$)

20___	20___	20___

4	Is the research topic in the area of priority within the strategic plan for health research in the Western Pacific Region? [Preference would be given to application dealing with priority areas identified within this plan, especially on health policy development.]
	☐ Yes ☐ No

5	Is the research proposed in this application currently being submitted, totally or in part, elsewhere for possible support?
	If so, to which organization(s)?
	By which date is a decision expected?

6	Institutional and national ethical clearance
6.1	Institutional ethical clearance letter enclosed
	☐ Yes ☐ No
6.2	National ethical clearance document enclosed
	☐ Yes ☐ No

7	Approval of national Ministry of Health or National Medical Research Council (or equivalent body)
	National approval document enclosed
	☐ Yes ☐ No

8	Applicant's signature
	Date: Signature

9	Institutional endorsement
	Head of institution Title:
	Name (pls print): Date:
	Signature:

II. SHEETS FOR RESEARCH PROJECT DESCRIPTION

Title of project:

1. Statement of the problem

2. Relevance of the problem to national or local health objectives (biomedical, beavioural and health systems development)

3. Field(s) of application of the proposed research results

4. Review of literature and other existing information

5. Statement of objectives

6. Variables

7. Statement of research hypotheses, if any

8. Research methodology

    a. Summary of methodology (not more than 150 words)

    b. Research design

        (1) Selection of research strategies

        (2) Selection of research setting

    c. Sampling

    d. Use of controls

    e. Study instrument(s)

    f. Short description of plans for collecting data

    g. Short description of plans for analysis of data and interpretation of results

9. Budget (use attached budgeting sheet)

III. CURRICULUM VITAE OF APPLICANT

1	Family name (surname) First name Other names
2	Date and place of birth
3	Nationality at birth at present
4	Academic qualifications and dates
5	Posts held (type of post, institution/authority, dates chronologically starting with present appointment)
6	Publications

(use extra pages if necessary)

IV BUDGET

1	PERSONNEL (allowance to be paid)	% or working time devoted to project	Year 1 US$	Year 2 US$	Year 3 US$	Total US$
1.1	Professional scientific staff (name and functional title)					
1.2	Technical staff (name and functional title)					
1.3	Other staff (name and functional title)					
	Subtotal					
2	MAJOR EQUIPMENT (more than US$500) (include specifications, shipment and freight insurance costs; comment on local provision for maintenance/service)		Year 1 US$	Year 2 US$	Year 3 US$	Total US$

			Year 1 US$	Year 2 US$	Year 3 US$	Total US$
	Subtotal					
3	SUPPLIES					
3.1	Chemicals					
3.2	Glassware					
3.3	Minor equipment (less than US$500 each) (include shipment and freight insurance costs)					
3.4	Animals					
3.5	Other supplies					
3.6	Operating costs (specify: maintenance of equipment, gasoline, etc.)					
	Subtotal					
4	TRAVEL (specify domestic and international		Year 1 US$	Year 2 US$	Year 3 US$	Total US$
	Subtotal					
5	DATA ANALYSIS COSTS					
	Subtotal					
6	MISCELLANEOUS EXPENDITURES					
	Subtotal					
7	SUMMARY		Year 1 US$	Year 2 US$	Year 3 US$	Total US$
	Personnel (1)					
	Major equipment (2)					
	Supplies (3)					
	Travel (4)					
	Data analysis cost (5)					
	Miscellaneous expenses (6)					
	GRAND TOTAL					

References and further reading

Brownlee A., Nchinda T., Mousseau-Gershman Y. Health services research course. Boston, MA, Boston University Medical Center, 1983.
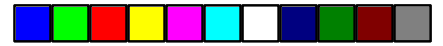
Chi I.C. et al. Vasectomy and non-fatal acute myocardial infarction: a hospital-based case-control study in Seoul, Korea. International Journal of Epidemiology, 1990, 19(1): 32-41.

Cushieri A., Baker P.R. Introduction to research in medical sciences. Edinburgh, Churchill Livingstone, 1977.

Marks R. Designing a research project. Belmont, CA, Lifetime Learning Publications, 1982.

Omran A.R. The Clark-Omran system of research design in epidemiology (paper used at the National Workshop on Research Design and Methodology on Biomedical Research, Manila, August 1986, supported by WHO/WPRO).

Annexes
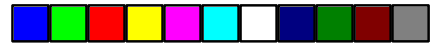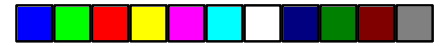
Annex 1

Questionnaire Design

Introduction

The use of questionnaires and interviews is a standard method of data collection in clinical, epidemiological, psychosocial and demographic research. Physicians are paramount practitioners of questionnaires and interviews in everyday practice: taking a medical history is a form of questionnaire interview and is recorded either in a fixed protocol (medical record) or taken as an open-ended interview.

Synonyms: schedules, inventories, study instruments
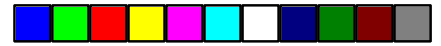
Definition: A questionnaire is simply a list of mimeographed or printed questions that is completed by or for a respondent. An interview schedule is a list of more or less structured questions that are read out or verbalized by an interviewer (with or without probing) in interrogating a respondent. The interviewer then records the respondent's replies either verbatim (for open-ended questions) or according to prespecified (or even precoded) answers or categories thereof.

1. Types

Mailed questionnaires require a literate respondent and, despite their low cost, usually result in a high rate of non-response. Further, there is no guarantee that respondents are no different from non-respondents (they usually are different).

Telephone interviews are easy to conduct in urban areas, but miss those without a telephone or those at work. They are of little use in developing countries, except for very selective studies.

Face-to-face interviews by a trained interviewer are the commonest form in community surveys and clinical research.

Standard questionnaires or inventories are specially prepared questionnaires, which are used with specified methods of scoring and analysis. Examples: Cornell medical index schedule, health opinion survey schedule, world fertility survey schedule, psychological tests and inventories such as the Minnesota Multiphasic Personality Inventory, London School of Hygiene cardiovascular questionnaire.

The latter is given in its entirety in Appendix 1. It is useful in community surveys of cardiovascular disease by trained interviewers (not physicians). You will notice that the answers are used to 'diagnose' the specified clinical conditions (angina, pain of possible infarction, intermittent claudication) by scoring answers to specific questions. In most situations, investigators write their own questionnaire.

2. Aims

The face-to-face format allows:

a. clarification of questions;

b. probing for answers (if allowed);

c. use of visual aids;

d. high response rate; and

e. short time in filling out the questionnaire.

However, it is expensive, requires training of interviewers and introduces interviewers' bias. It is difficult to recruit professional interviewers.

The information solicited by questionnaires may be:

a. facts, such as age or disease;

b. knowledge about, e.g. services and programmes;

c. attitudes/opinions, e.g. about contraception, immunization and breast-feeding;

d. behaviour, such as use of health services, traditional medicine, prenatal care, dental hygiene, periodic check-ups and smoking;

e. compliance with doctor's instructions; and
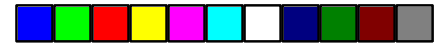
f.   information about others (research gossip), e.g. a mother is asked about her children or her husband; neighbours are asked about an event in the vicinity, or whether women still use the <u>daya</u> (traditional birth attendant).

3. Factors

Factors to be considered in the design include:

a.   study objectives and major research questions;

b.   study hypotheses: what data are required to accept or reject a hypothesis;

c.   data to be collected;

d.   plans for analysis and dummy tables so that no important information is missed;

e.   budget; and

f.   the audience or target population: age, sex, religion, language, traditionalism, stranger in the house (can a wife be interviewed in the absence of the husband?); above all, will the respondents be able to give the required answers?

Questionnaire format

1.   Open and closed questions
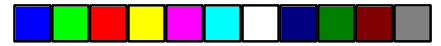
Structured closed questionnaires have the advantages of being:

-    focused and pertinent to the study objectives,

-    easy to administer,

-    uniform,

-    precoded and thus easy to analyse, and

-    analysed in a short time.

They are preferred in medical studies.

Open-ended questionnaires are useful for anthropological and social enquiries. Some of the questions in medical surveys may be open-ended, but the fewer the better. Such questions allow the respondent to talk freely and at length, but he may deviate from the

subject in question. They require special coding after the end of the study, thus lengthening the time for analysis.

2. *Forms of structured questions*

Structured questions may offer:

a. a dichotomous choice of 'yes' or 'no', 'approve' or 'disapprove', or 'effective' or 'not effective'. Questions of this type should always include a 'don't know' response category.

b. multiple choice of items:

Example: To whom do you go first for advice on contraceptive methods?

- the <u>daya</u>;

- your friends or neighbours;

- your mother-in-law;

- the nurse;

- the doctor;

- the pharmacist;

- others (specify).

c. A rating scale such as the multiple-step scale in semantic differential requires the respondents to grade their answers between two extremes (e.g. 'bad' <u>versus</u> 'good'; 'approve' <u>versus</u> 'disapprove').

Example: Bad   1 2 3 4 5 6 7   Good

Approve 1 2 3 4 5   Disapprove

Such extended ranges are artefactual and are certainly beyond the comprehension of many people.

d. a numerical answer, e.g. 'How old are you?'

e. an opportunity for probing to elicit more specific responses. Probing is sometimes allowed, but training of interviewers is essential to maintain the uniformity of this approach.
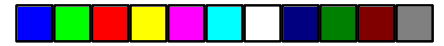
3. Language and wording style

The language of the questions should be pitched to the level of the respondent. A common, everyday, conversational style or vernacular should be used. In cross-cultural studies, questionnaires are translated from the original language into the local language or dialect, say from English to Swahili. They are then translated back to English by an independent linguist to check and correct any possible misunderstanding.

Avoid leading questiond, e.g. 'Don't you think that the intrauterine device is safer than the pill?' It would be better to ask: 'Which do you think is safer, the intrauterine device or the pill?'

Avoid professional jargon and abbreviations.

4. Coding responses to questions

The response categories should include all possible responses. This usually means including a 'don't know' or 'sometimes' or 'maybe' category. Much time can be saved in the analysis if responses are scaled at the same time as they are recorded. For example, responses to the question: 'Is there a physician present when you visit the clinic?' could be coded:
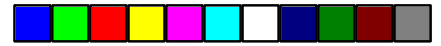
Always	4	Usually	3	Sometimes	2
Don't know	1	Never	0		

The response categories must be mutually exclusive, that is, the answer categories must not overlap.

There should be no blank left in any coding space, since a blank could mean that the question was either inadvertently omitted or purposely skipped. For this reason, all questions that may be skipped by certain categories of respondents (e.g. ages of children for childless couples) should be provided with an 'NA' (not applicable) code, usually 9 or 99.

Some data can best be obtained through a 'cascade' format. In the example shown in Figure 1, note how several questions have been used to scale a response to the unasked but real underlying question, 'How much new information have you learn it from family planning posters?' Notice that all responses are treated as the response to a single question and coded in only one location (rather than using three different code columns for the three questions).

FIGURE 1 EXAMPLE OF A 'CASCADE' FORMAT OF QUESTIONNAIRE

Have you ever seen a sign or poster for family planning anywhere?

No   1   GO TO QUESTION 10

Yes   The last time you saw such a poster did you
read the information on it?

            No   2   GO TO QUESTION 10

            Yes   Did the poster give you new
information about family
planning that you didn't
already know?

                        No   3

                        Don't know   4

                        Yes   5

5. *Length of questionnaire*

Most survey experts agree that a questionnaire of reasonable length is one that would take half an hour or less to administer in a structured interview. Stretching this time to three quarters of an hour is frowned upon, while longer questionnaires are considered counterproductive. It is true that a trained interviewer may keep the interest of a respondent for an hour, especially in an open-ended interview, but this is rare and expensive. Most interviews today are conducted by semi-professional interviewers with *ad hoc* training. Few of them will be able to conduct lengthy interviews without fatigue setting in and biasing the responses.

6. *Reliability of questionnaires*

Two means of ensuring reliability can be used:

a.   in-built reliability, which is achieved by repeating certain questions, rephrasing the second inquiry while maintaining the same or comparable response codes; and

b.   repeat reliability, which is achieved by repeating the interview with a small percentage of the respondents (chosen at random). Usually, factual questions are used to measure reliability: opinion questions do not provide
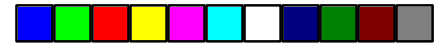
a direct measure of reliability, because people change their minds from time to time. Revision of opinion may be the subject of special methodological studies, however.

7. *Validity/consistency checks*

Certain items in a questionnaire may be validated in special surveys. For example, clinical records can be checked against the responses of women who have been receiving injectables from the clinic regularly over the preceding 12 months.

Another type of validity check is the consistency or cross-check. If a woman is 18 years old, she cannot possibly have a child aged 10 or 15; a woman with two single pregnancies cannot have three infant deaths. This tedious job can be done by computer.

8. *Layout of questionnaire*
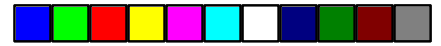
The layout of the questionnaire should be physically pleasant and artistically tasteful. Questionnaires should not become display items, however, and budget considerations have to be seriously considered. Money needlessly put into production of a questionnaire will not be available for the survey itself.

9. *Sequencing of questions*

The questions should be asked in a proper sequence. The following are some general guidelines:

a. Introduction: A clear and concise but relevant introduction to the questionnaire is helpful. It should seek to identify the investigator or interviewer with a respected agency in the community. It should indicate the purpose of the questionnaire and should remove any hesitation on the part of the respondent. Sometimes an identity card is essential.

b. Cover sheet or identification page (see Figure 2). This page usually carries:

- the name of the survey and the responsible organization;

- the code for the respondent or household; and

- the name of the interviewer and date of the interview.

> c. Warm-up questions or statements should start the questionnaire itself. Do not start with threatening questions about income and other sensitive issues.
>
> d. The transition from one section to the other should be smooth.
>
> e. In the body of the questionnaire, appropriate use should be made of standard formats for instructions: boxes for instructions, and arrows for directions and which directions to skip (instructions for questions that should be bypassed for a particular respondent; see Figure 3).
>
> f. Instructions: Two possibilities can be envisaged: (i) a separate instruction manual may be used, or (ii) instructions may be included in the questionnaire itself. In such cases, they should be distinguished from questions by putting them in boxes, or writing them in capital letters or italics or other special typefaces (see Figure 4).

Auxiliary activities

1. Pretesting the questionnaire

A pretest is a try-out of the questionnaire. Pretesting is carried out on a small number of respondents who are comparable with the sample of correspondents but are not part of it. The results of pretesting are incorporated into the rewriting of the questionnaire. Even if a standardized questionnaire is used, it should be pretested in the population being studied, and a reliability coefficient calculated.

2. Training of interviewers

Interviewers must be carefully selected and properly trained. In survey research, they become the backbone of data collection. It would be false economy to 'economize' on interviewers and go to great expense for other aspects of the study. Role-playing is as essential as is supervision.

Instructions should be given about confidentiality of information, patience and perseverance, being pleasant, with a positive attitude, following instructions, etc. Interviewers should always be supervised (one supervisor to four to six interviewers).

3.   Call-backs

Call-backs or repeat visits to non-respondents are most helpful in minimizing the non-response rate. The time of the call-back should coincide with the time that the respondent is most likely to be home. Persons who have refused to participate should also be revisited in the hope that they may cooperate. Call-backs add, however, to the cost of a survey and there must be a limit on how many can be done. Perhaps two or three call-backs to non-respondent are enough.

4.   Editing and coding

Questionnaires should be checked by supervisors at the end of each day for omissions, incomplete answers, unclear statements or illegible writing. Interviewers may have to go back to collect missing or unclear information. Responses are then carefully coded, with verification.

FIGURE 2 AN EXAMPLE OF A COVER SHEET FOR A QUESTIONNAIRE

Country _____

Household schedule   Area/Zone: _____

Identification number

1	2	3	4	5	6	7	8	9	10	11	12

Columns 1-2   :   Country _____
(Code last)   Area _____
                           Religion (P.11) _____
                           Social class (P.35, EW) _____

Columns 3-4   :   Card Number 01 _____

Columns 5-8   :   Household serial number _____

Columns 9-10   :   Total households size (P.11) _____

Column 11   :   Total no. eligible women (P.11) _____

Column 12   :   Type of family (P.11) _____

Address of household   :_____

                           _____

Household number   :_____

Name of head of household : _____

Name of respondent (if not head of household) and relation to

☐ head of household   :_____

Name of interviewer   :_____

Date of interview   : _____

IBM COL: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

PREGNANCY HISTORY CHART

QUESTION: a-1	IBM CARD NO.	PREG-NANCY ORDER	a. MAT. AGE AT TERMI-NATION	b. PREGNANCY OUTCOME	GESTATION PERIOD	c. BIRTH INTERVAL	d. PLACE/ ATENDANT AT PREGN. TERMINATION	LIVE BIRTH ORDER	e. SEX OF CHILD	f. CHILD STILL LIVING	g. AGE OF LIVING CHILD	h. AGE AT DEATH	i. DAUGHTER'S AGE AT MENSES
INSTRUCTIONS FOR CODING ▮▮▮▮▮ WRITE NAMES OF ALL LIVE BIRTHS BELOW: ↓		CODE 01 FOR FIRST AND THEN 02. 03 ETC. FOR EACH PREG-NANCY	RECORD ACTUAL AGE IN YEARS (PROBE)	CODE IN COL. 20: 1 - Live Birth 2 - Stillbirth ABORTION 3 - Spontaneous 4 - Induced 5 - Unspecified CODE GESTATION PERIOD IN WEEKS FOR EACH PREGNANCY IN COLS. 21-22 Code 55 - Don't Remember		CODE NO. OF MONTHS SINCE TERMI-NATION OF LAST PREG-NANCY CODE 97-97 months	CODE: 1- Hosp/ Medical 2- Hosp/Para-medical 3- Home/ Medical 4- Home/Para-medical 5- Home/Non-medical 6- Home/ Alone		CODE: 1 Male 2 Female	CODE: 1 Living 2 Dead	IF AGE IS/WAS ONE YEAR OR OLDER, RECORD ACTUAL NUMBER IF LESS THAN 1 YEAR, CODE: 77- 6-12 months 88- 1-6 months 99- less than one month		RECORD AGE WHEN SHE STARTED HER PERIODS CODE: 77- Male child 88- Has not yet started
IBM COL ▮▮▮▮	3-4	16-17	18-19	20	21-22	23-24	25	26-27	28	29	30-31	32-33	34-35
	05												
	06												
	07												
	08												
	09												
	10												
	11												
	12												
	13												
	14												
	15												
	16												
	17												
	18												
	19												

PARITY: TOTAL NO. OF LIVE BIRTHS (CODED "I") IN COL _____   FAMILY SIZE: TOTAL NO. OF LIVING CHILDREN (CODED "I" IN COL. 29) _____

FIGURE 3   USE OF BOXES AND ARROWS IN QUESTIONNAIRES

FIGURE 4   EXAMPLE OF THE USE OF CAPITAL LETTERS DISTINGUISHING INSTRUCTIONS

19. Do you feel that your last pregnancy came sooner than you wanted?   46

    1-Yes   8-No   9-No answer

    IBM CARD 02 CONTINUED AFTER PREGNANCY HISTORY

BEGIN SEPARATE CARD FOR EACH PREGNANCY

---

20. I am going to take down the details of your pregnancies one by one. Let us begin with your first pregnancy.

    RECORD ANSWERS TO QUESTIONS 20 a-l ON PREGNANCY HISTORY CHART, USING NUMERALS GIVEN IN CODING INSTRUCTIONS. DO NOT WRITE WORDS ON THE CHART EXCEPT TO RECORD NAMES OF ALL LIVE BIRTHS: ASK QUESTIONS ABOUT EACH PREGNANCY. BEGIN WITH FIRST PREGNANCY AND CONTINUE THROUGH TO MOST RECENT PREGNANCY. USE BACK OF PAGE FOR ADDITIONAL PREGNANCIES.

---

a. What was your age when you terminated this pregnancy? (PROBE) RECORD ACTUAL NUMBER.

---

b. Was the outcome of this pregnancy a live birth, a stillbirth (fetal death), a multiple birth, or an abortion?

    How many weeks had you been pregnant when the pregnancy terminated? (RECORDGESTATION PERIOD IN WEEKS IN COLS. 21-22)

    INTERVIEWER MAY DEFINE 'STILLBIRTH' AS MISCARRIAGE OVER SEVEN MONTHS; DEFINE 'ABORTION' AS MISCARRIAGE UNDER SEVEN MONTHS.

    IF RESPONSE IS 'MULTIPLE BIRTH', TREAT EACH FETUS SEPARATELY (ASK QUESTIONS d-1 ABOUT EACH AND RECORD ANSWERS ON SEPARATE LINES OF CHART.) THE INFORMATION ON PREGNANCY ORDER, MATERNAL AGE, GESTATION PERIOD AND BIRTH INTERVAL WILL BE THE SAME FOR EACH FETUS OF MULTIPLE BIRTH.

    IF RESPONSE IS 'ABORTION', ASK THE FOLLOWING QUESTION: Was this abortion induced (did you cause it yourself or have someone else cause it OR was it spontaneous (miscarriage)?

---

c. How many months had it been since you terminated your last pregnancy? FOR FIRST PREGNANCY, CALCULATE INTERVAL SINCE MARRIAGE OR CONSUMMATION (WHICHEVER IS APPLICABLE).

    FOR SUCCEEDING PREGNANCIES, CALCULATE INTERVAL IN MONTHS BETWEEN LAST TERMINATION AND END OF THIS PREGNANCY. RECORD ACTUAL NUMBER OF MONTHS IN INTERVAL.

---

d. Who attended you and where was this pregnancy terminated?

    HOSPITAL: INCLUDE CLINIC AND HEALTH CENTRE

    HOME: INCLUDE PLACES OUTSIDE HOSPITAL

CODER: ENUMERATE LIVE BIRTH ORDER CONSECUTIVELY (01,02,03, etc.) IN COLUMNS 26-27 OF PREGNANCY HISTORY CHART.
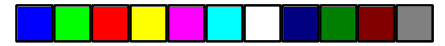
Appendix 1

LONDON SCHOOL OF HYGIENE CARDIOVASCULAR QUESTIONNAIRE
(FOR ADMINISTRATION BY AN INTERVIEWER)

Section A: Chest pain on effort

1. Have you ever had any pain or discomfort in your chest?

    ____   Yes

    ____   No

    If 'No', proceed to Section C.

    If 'Yes', ask next question. (if, during the remainder of Section A, an answer is recorded in a box marked; proceed to Section B.)

2. Do you get it when you walk uphill or hurry?

    ____   Yes

    ____   No

    ____   Never hurries or walks uphill

3. Do you get it when you walk at an ordinary pace on the level?

    ____   Yes

    ____   No

4. What do you do if you get it while walking?

    ____   Stop or slow down

    ____   Carry on

    Record 'stop or slow down' if subject carries on after taking nitroglycerine.

5. If you stand still, what happens to it?

    ____   Relieved

    ____   Not relieved

6. How soon?

    ____   10 minutes or less

    ____   More than 10 minutes

7. Will you show me where it was?

    _____ Sternum (upper or middle)

    _____ Sternum (lower)

    _____ Left anterior chest

    _____ Left arm

    _____ Other

    (Record all areas mentioned)

8. Do you feel it anywhere else?

    _____ Yes

    _____ No

    (If 'Yes', record additional information above)

Section B: Possible infarction

9. Have you ever had a severe pain across the front of your chest lasting for half an hour or more?

    _____ Yes

    _____ No

Section C: Intermittent claudication

If an answer is recorded in a box marked; no further questions need to be asked.

10. Do you get pain in either leg on walking?

    _____ Yes

    _____ No

11. Does this pain ever begin when you are standing still or sitting?

    _____ Yes

    _____ No

12. In what part of your leg do you feel it?

    _____ Pain includes calf/calves

    _____ Pain does not include calf/calves

    If calves not mentioned, ask: Anywhere else?

13. Do you get it if you walk uphill or hurry?

    _____   Yes

    _____   No

Never hurries or walks uphill

14. Do you get it if you walk at an ordinary pace on the level?

    _____   Yes

    _____   No

15. Does the pain ever disappear while you are walking?

    _____   Yes

    _____   No

16. What do you do if you get it when you are walking?

    _____   Stop or slow down

    _____   Carry on

17. What happens to it if you stand still?

    _____   Relieved

    _____   Not relieved

18. How soon?

    _____   10 minutes or less

    _____   More than 10 minutes

<center>Diagnostic criteria for angina pectoris,
possible infarction and intermittent claudication</center>

'Angina' is defined as being present in subjects who answer as follows:

Q.1	: 'Yes'
Q.2 or 3	: 'Yes'
Q.4	: 'Stop or slow down'
Q.5	: 'Relieved'
Q.6	: '10 minutes or less'
Q.7	: (a) Sternum (upper or middle, or lower), or (b) left anterior chest and left arm

(If interviewing instructions are correctly observed throughout, it is sufficient to check the answer to Q.7.)

'Angina' may be graded according to severity:

      Q.3   : 'No' = Grade 1

                         'Yes' = Grade 2

'Pain of possible infarction' is defined as being present in subjects who answer as follows:

      Q.9   : 'Yes'

'Intermittent claudication' is defined as being present in subjects who answer as follows:

      Q.10   : 'Yes'

      Q.11   : 'No'

      Q.12   : 'Includes calf'

      Q.13 or 14   : 'Yes'

      Q.15   : 'No'

      Q.16   : 'Stop or slow down'

      Q.17   : 'Relieved'

      Q.18   : '10 minutes or less'

'Intermittent claudication' may be graded according to severity:

      Q.14   : 'No'   = Grade 1

                         'Yes' = Grade 2

Annex 2

Method for Presenting and Interpreting Health-related Data: Tables, Graphs and Charts

I. Tables

Although there are no hard and fast rules governing table construction, there are certain general principles that have become accepted as more or less standard.

A. Tables should be as simple as possible. Two or three small tables are preferred to a single large table containing many details or variables. Generally, three variables constitute a maximum number which can be read with ease.

B. Tables should be self-explanatory.

   1. Codes, abbreviations or symbols should be explained in detail in a footnote.

   2. Each row and each column should be labelled concisely and clearly.

   3. The specific units of measure for the data should be given.

   4. The title should be clear, concise, and to the point. Answers: what? when? where?

   5. Total should be shown.

C. The title is commonly separated from the body of the table by lines or spaces. In small tables, vertical lines separating the columns may not be necessary.

D. If the data are not original, their source should be given in a footnote.

E. Specific examples

1. The simplest table is a two-column frequency table. The first column lists the classes into which the data are grouped. The second column lists the frequencies for each classification. An example is shown in Table 1.

**TABLE 1: CLASSIFICATION OF LIVE BIRTHS
BY EDUCATION OF FATHER, ANY STATE, 1968**

Education of father	Number of live births
High school graduate	50 684
Less than 12 years of school	31 774
TOTAL	82 458

Source: "Vital Statistics of the United States", 1968, Volume 1, p.81

2. Table 1 may be enlarged to included subclassification, such as place of delivery and attendant at birth, as shown in Table 2.

**TABLE 2: CLASSIFICATION OF LIVE BIRTHS BY EDUCATION OF FATHER,
PLACE OF DELIVERY AND ATTENDANT AT BIRTH, ANY STATE, 1968**

Education of father	NUMBER ATTENDED BY				Total
	Physician		Midwife	Others and not specified	
	In hospital	Not in hospital			
High school graduate	46 606	3 014	910	154	50 684
Less than 12 years of school	14 334	3 094	13 930	416	31 774
Total	60 940	6 108	14 840	570	82 458

Source: 'Vital Statistics of the United States', 1968, Volume 1, p.81

3. Summarization of data will be expedited and simplified by initially preparing a master table. In this table, all available data should be completely classified. When complete cross-classification is made, data relative to a single variable or to any combination of variables may be obtained without recourse to the original data.

From the general format of a master table as shown in Table 3, we can determine how many URBAN PERSONS (A), how many MALES (B), and how many URBAN MALES IN A SPECIFIED AGE GROUP (C), were admitted to the hospital, etc.

TABLE 3. ADMISSION TO ANY HOSPITAL FOR THE YEAR 1968, CLASSIFIED BY AGE, RESIDENCES, AND SEX

Age in years	Urban			Rural			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
<1									
1-4									
5-9									
10-14	C								
.									
.									
.									
Total			A				B		

II. Graphs

Definition: A graph is a method of showing quantitative data using a coordinate system (for our purposes, usually x and y).

There are several different types of graphs, a few of which are rectangular coordinates, polar coordinates, special purposes types (3-dimensional), etc. We will limit ourselves to rectangular coordinate graphs.

A.   General concept

Rectangular coordinate graphs are those which consist of two sets of lines which are at right angles to each other. On each set of lines there is a scale of measurement for the purpose of identification. Figure 1 presents the general structure of rectangular coordinate graphs. Generally the variable assigned to the x-axis in considered the independent variable (method of classification), whereas the variable assigned to the y-axis is the dependent variable (frequency). That is, in drawing a graph, we plot a change in 'y' with respect to x.

FIGURE 1: GENERAL GRAPH

METHOD OF CLASSIFICATION
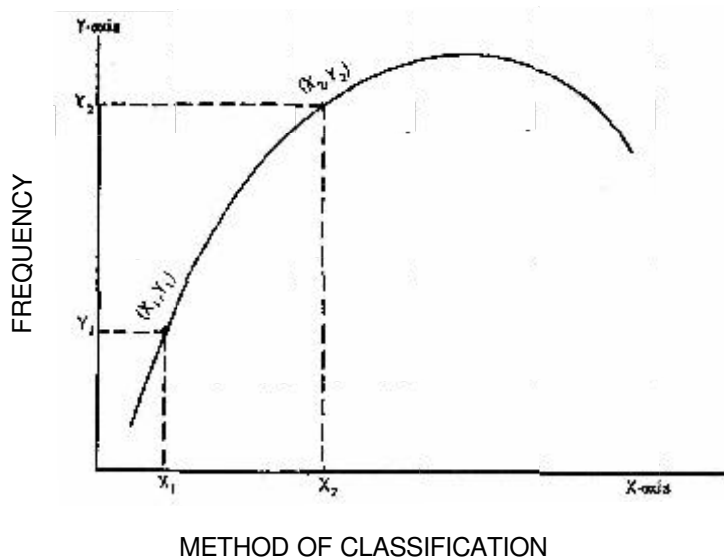
B. General principles

When graphs have been drawn correctly, they allow the reader rapidly to obtain an overall grasp of the data. Some of the most important principles of graphing are:

1.   The simplest graphs are the most effective. No more lines or symbols should be used in a single graph than the eye can easily follow.

2.   Every graph should be self-explanatory.

3.   The title may be placed either at the top or bottom of the graph.

4.   When more than one variable is shown on a graph, each should be clearly differentiated by means of legends or keys.

5.   No more coordinate lines should be shown than are necessary to guide the eye.

6.   Lines of the graph itself should be <u>heavier</u> than other coordinate lines.

7.   Frequency is usually represented on the vertical scale and method of classification on the horizontal scale.

8.   On an arithmetic scale, equal increments on the scale must represent equal numerical units.

9.   Scale divisions should be clearly indicated as well as the units into which the scale is divided.

C.   Specific examples

1.   Arithmetic scale line graph

A scale line graph is one where an equal distance represents an equal quantity anywhere on the axis, but not necessarily between the axes. Care must be exercised in the choice of whether we use equal intervals on both axes, wide intervals on the x-axis in relation to the y-axis or vice versa. The scales should be defined in such a way that the final product is pleasing to the eye. A scale break may be used with a scale line graph, but if used, care must be taken lest the graph be misinterpreted. Figure 2 is an example of a scale line graph.

**FIGURE 2: REPORTED RUBELLA CASE RATES
BY 13 FOUR-WEEK PERIODS, UNITED STATES 1968**

PERIOD ENDING
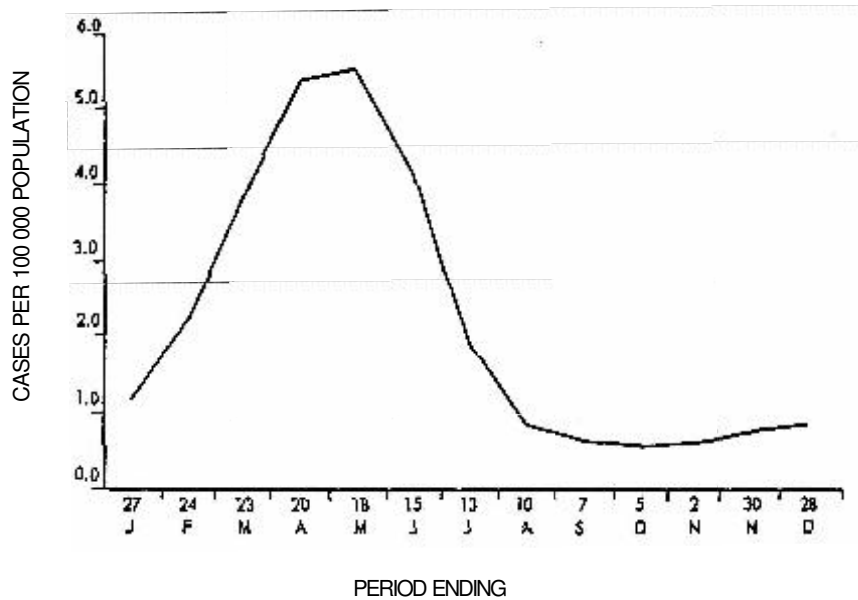
Source: Rubella Surveillance Unit

2.   Semi-logarithmic scale line graph

The semi-logarithmic scale line graph is one where one coordinate or axis, usually the y-axis, is measured in logarithms of units, whereas the other axis is measured in arithmetimc units. This is useful in that when examining a series of data over a period of tie we are often interested in the relative (or rate of) change rather than the absolute (actual amount of). The advantages of semilog graphing are:

a)   a straight line indicates a constant rate of change,

b)   the slope of the line indicates the rate of increase or decrease,

c)   two or more lines following parallel paths show identical rates of increase (or decrease).

An illustration of this type of graph is shown in Figure 3.

**FIGURE 3: REPORTED ANNUAL INCIDENCE RATES, DEATH RATES
AND DEATH-TO-CASE RATIOS FOR DIPHTHERIA, UNITED STATES, 1920-1968**
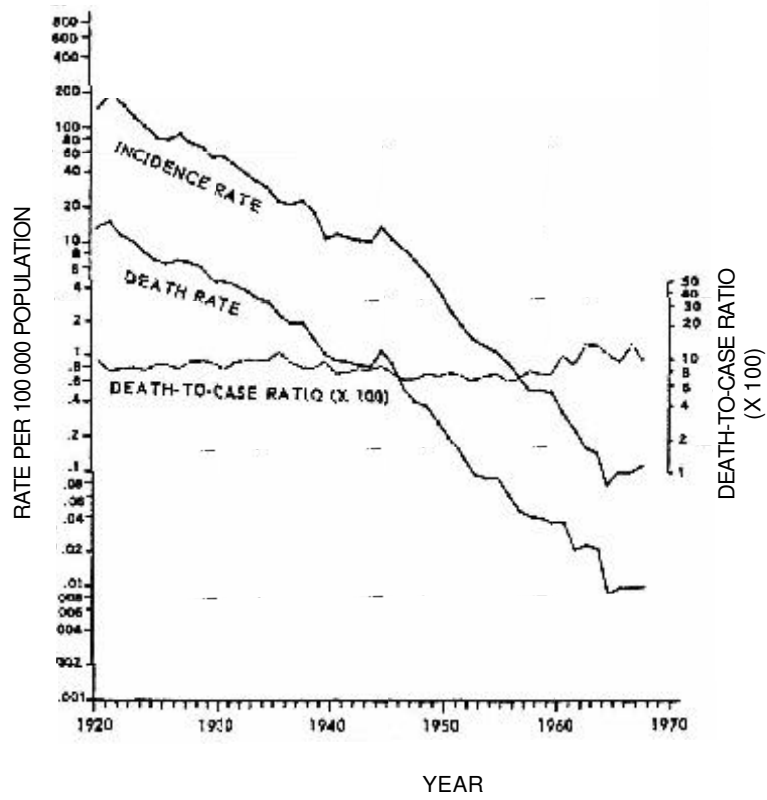
Source: Diphtheria Surveillance Unit

3.   Histogram

A histogram is a graph used only for presenting frequency distribution of quantitative data. There is no space between the cells (often referred to as tic-marks) on a histogram. This graph is not to be confused with a bar chart which has space between the cells. A scale break should not be used in the histogram because the histogram depicts the total area under the curve. Because of this characteristic the easiest type of histogram to construct will be one of equal class intervals as shown in Figure 4.

FIGURE 4: CASES OF RASH ILLNESS, ELEMENTARY SCHOOL
SAMPLE CITY, FEBRUARY 22 - MARCH 23, 1970
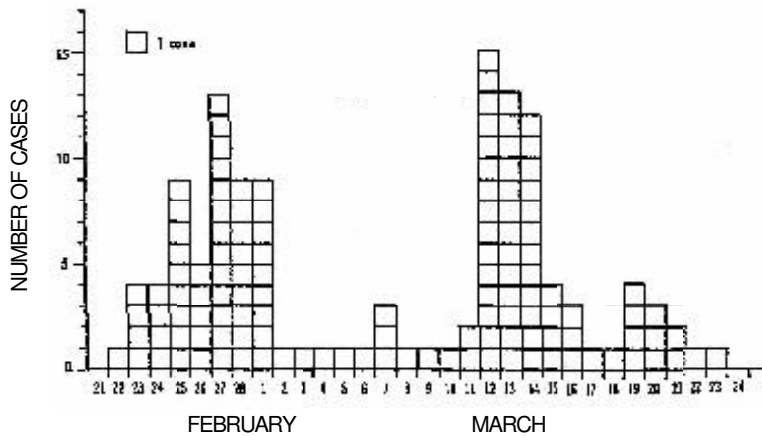
For illustration, Figure 4 shows the area under the curve partitioned into each case of illness. Ordinarily, only the line representing the height of each column would be drawn.

In order that the area of each rectangle in the histogram represents a specified number of cases, we let the height represent the number of cases per unit of measurement (in Figure 4, per day) and the width be the method of classification (in Figure 4, interval of time in days). Therefore, the height times the width will equal the number of cases within a day, just as the height times width is equal to the area of a rectangle.

A histogram with unequal intervals requires more thought in its construction because of the total area concept. Figure 5 has been selected to illustrate this situation.

**FIGURE 5: REPORTED CASES OF TETANUS
BY FIVE-YEAR AGE GROUPS, UNITED STATES 1968**
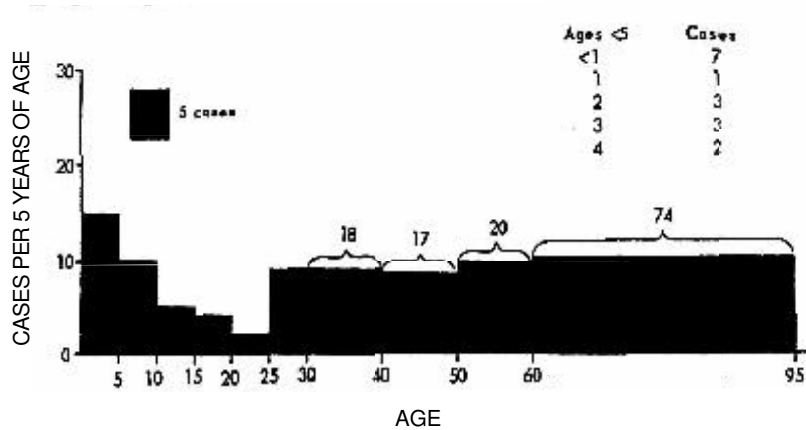
Source: MMWR Annual Supplement, 1968

In general, only one set of data should be shown on a histogram; however, it is quite common in the field of public health to present data for cases - deaths, males - females, etc., by the histogram.

4. Frequency polygon

If it is desired to present more than one set of data in terms of a frequency distribution, the data should be presented in the form of a frequency polygon. A frequency polygon is constructed from a histogram, i.e. the midpoints of the class intervals are connected by a straight line. This is illustrated in Figure 6.

**FIGURE 6: NUMBER OF CASES OF INFLUENZA-LIKE
ILLNESS BY WEEK, SAMPLE CITY, 1970**
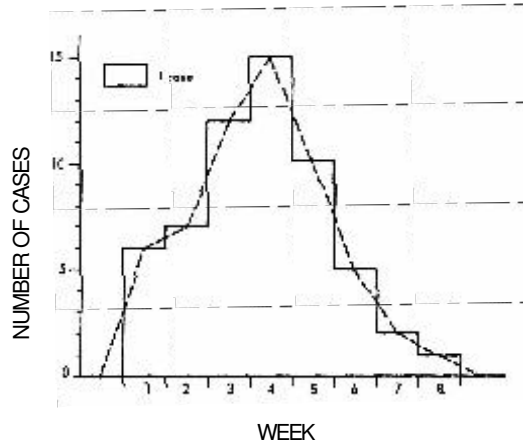
Since a frequency polygon is constructed from a histogram, the rules pertaining to area under the curve, equal and unequal class intervals will hold. Therefore, it is necessary to illustrate the manner in which this concept is preserved. The frequency polygon should be 'closed' in order to portray area. This 'closing' is done by connecting the first and last points with the base of the graph. The area in the frequency polygon must be approximately equal to that which would have been in the histogram. Figure 7 shows the <u>correct method</u> of closing the frequency polygon. Figure 8 shows the <u>incorrect method</u>.

**FIGURE 7: CORRECT METHOD
OF CLOSING FREQUENCY POLYGON**
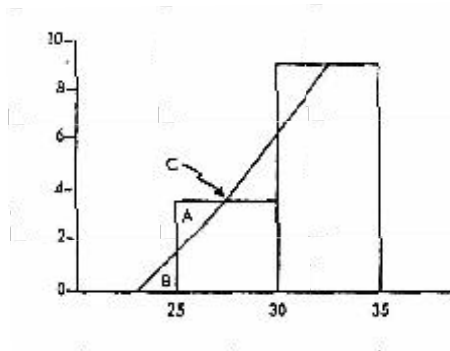
**FIGURE 8: INCORRECT METHOD
OF CLOSING FREQUENCY POLYGON**
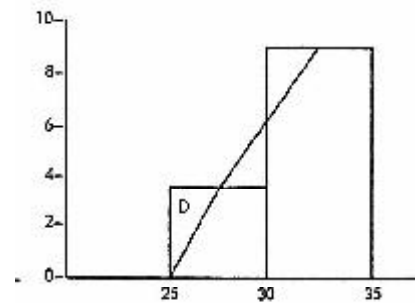
In Figure 7 the area designated by A would be part of the histogram if data were plotted by that method. In order to compensate for this area, which is excluded by the polygon, the point C is connected to the base line in such a manner that the area designated by B will be approximately equal to area A.

Figure 8 illustrates the incorrect method of closing a frequency polygon because the entire area designated by D is omitted and there is no provision for compensation.

In Figure 9 a frequency polygon having equal class intervals is shown. Additionally, Figure 10 illustrates a frequency polygon with unequal class intervals.

**FIGURE 9: NUMBER OF CASES OF INFLUENZA-LIKE ILLNESS
BY WEEK, SAMPLE CITY 1970**
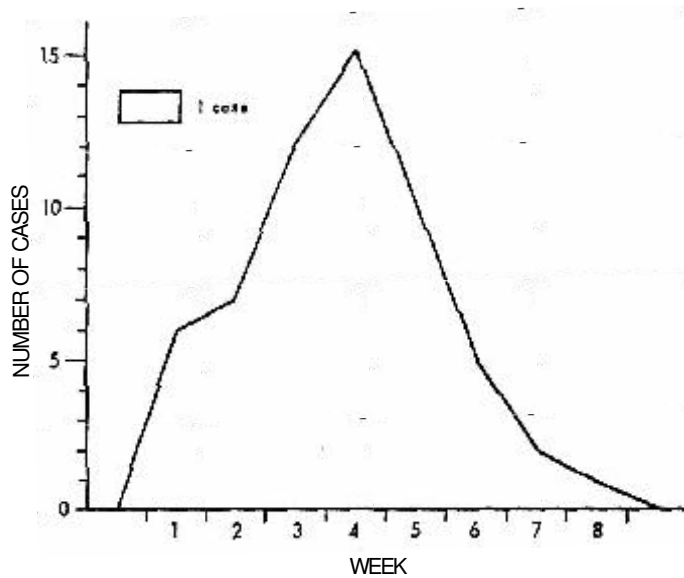
**FIGURE 10: REPORTED CASES OF TETANUS BY FIVE-YEAR AGE INTERVALS,
UNITED STATES, 1968**
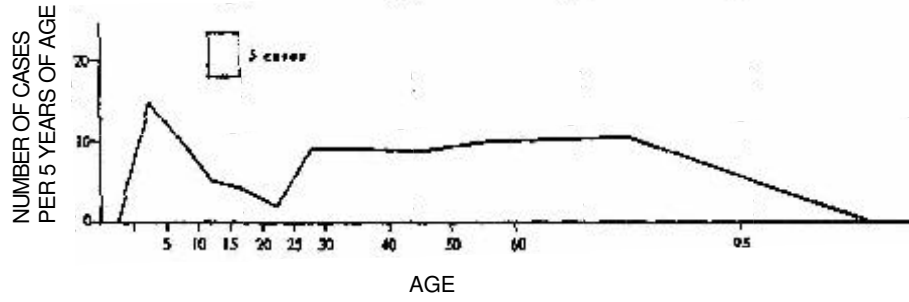
Source: MMWR Annual Supplement, 1968

A frequency polygon illustrating three sets of data is shown in Figure 11.

**FIGURE 11: REPORTED CASES OF ENCEPHALITIS BY MONTH
ACCORDING TO ETIOLOGY, UNITED STATES 1965**
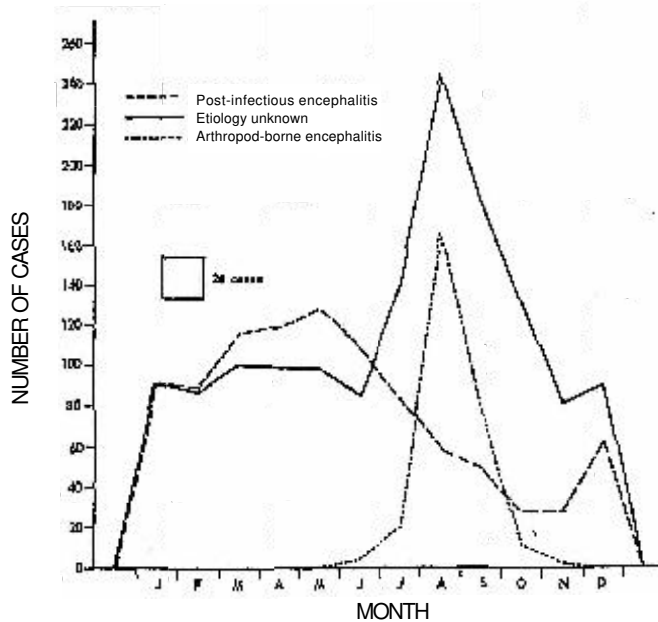
Source: 1965 Encephalitis Surveillance Report

5.   Scatter diagram

A scatter diagram is a special type of graph useful in pointing out relationships or associations between two variables. In this type of graph several sets of paired data are plotted on the same graph. The pattern made by the plotted points is indicative of a possible relationship. If they tend to follow a straight line, the relationship is of a linear nature. If the pattern does not follow a straight line, the relationship is curvilinear. If the pattern is just a scatter of points, then the trend suggests that probably no relationship exists. Figure 12 shows a scatter diagram.

**FIGURE 12: HISTOPLASMIN SKIN TESTS
COMPARISON OF OBSERVATIONS OF TWO READERS IN 51 SUBJECTS,
GREENWOOD, SOUTH CAROLINA, 1963**

*13 recorded as zero by both observers

*13 recorded as zero by both observers
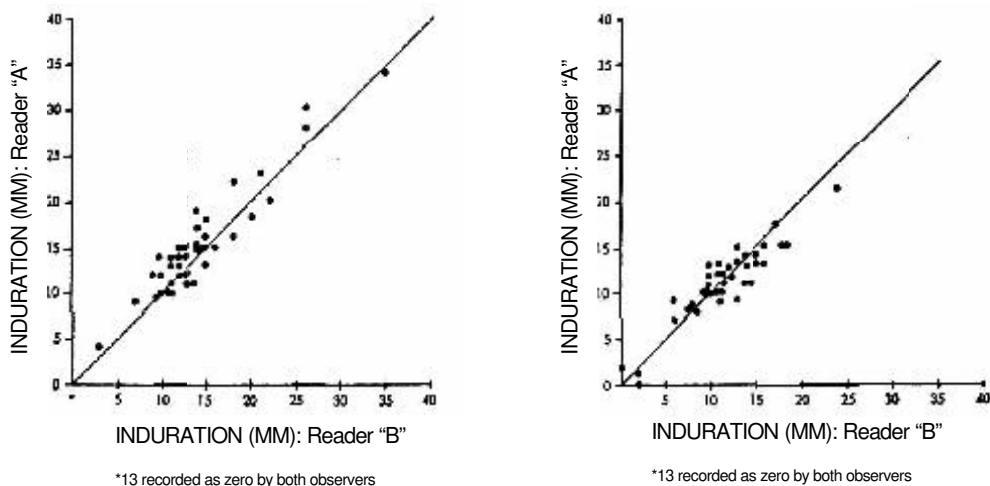
Source: Millar, J. D. et al: *American Review of Respiratory Disease*, Vol. 100, 1969

III. Charts

Definition: Charts are methods of presenting statistical information symbolically using only one coordinate.

There are many different types of charts -- some of those based on length, proportion, geographical coordinates, and a few special purpose charts will be presented in this section.

A. *Charts based on length*

Two of the most important charts under this category are bar charts and pictograms.

1.   Bar chart

The bar chart has cells, all of the same column width (unlike the histogram). There are also spaces between the columns (also unlike the histogram). This type of chart is ideally suited for presenting comparative data. The bars may be arranged horizontally as well as vertically (as illustrated in Figures 13 and 14). It is best to arrange these bars in either ascending or descending order for ease of reading. A scale break should never be used with a bar chart as this will lead to misinterpretation. Columns may be shaded, hatched or coloured, to emphasize differences between the bars. The bars should be labelled at the bottom and not in the middle of the chart itself, as this adds unnecessary confusion to a chart whose main virtue is its simplicity. When comparisons are made, the space between bars in the same group is optional, but space between groups is mandatory.

**FIGURE 13: PERCENT AGE DISTRIBUTION OF SAMPLE POPULATION
COMPARED TO PERCENT AGE DISTRIBUTION
OF CENSUS POPULATION IMMUNIZATION SURVEY, SAMPLE CITY, 1970**
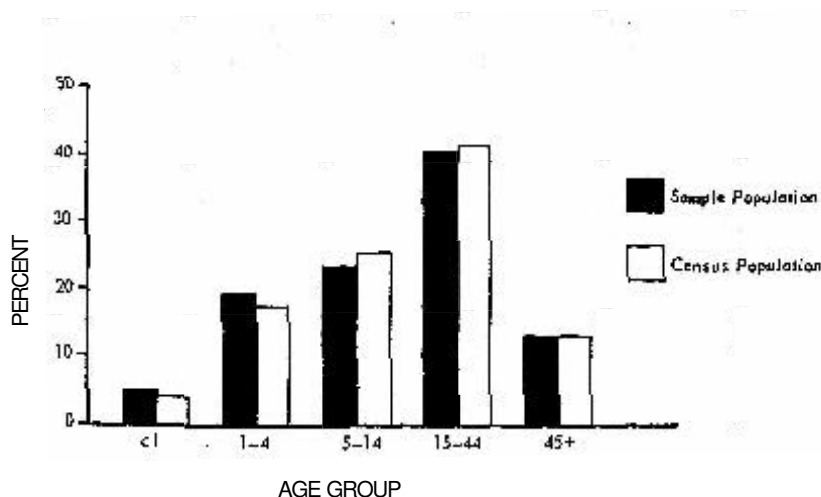
FIGURE 14: CRUDE DEATH RATE IN 20 COUNTRIES, 1955
(Rates are the number of deaths per 1000 population)
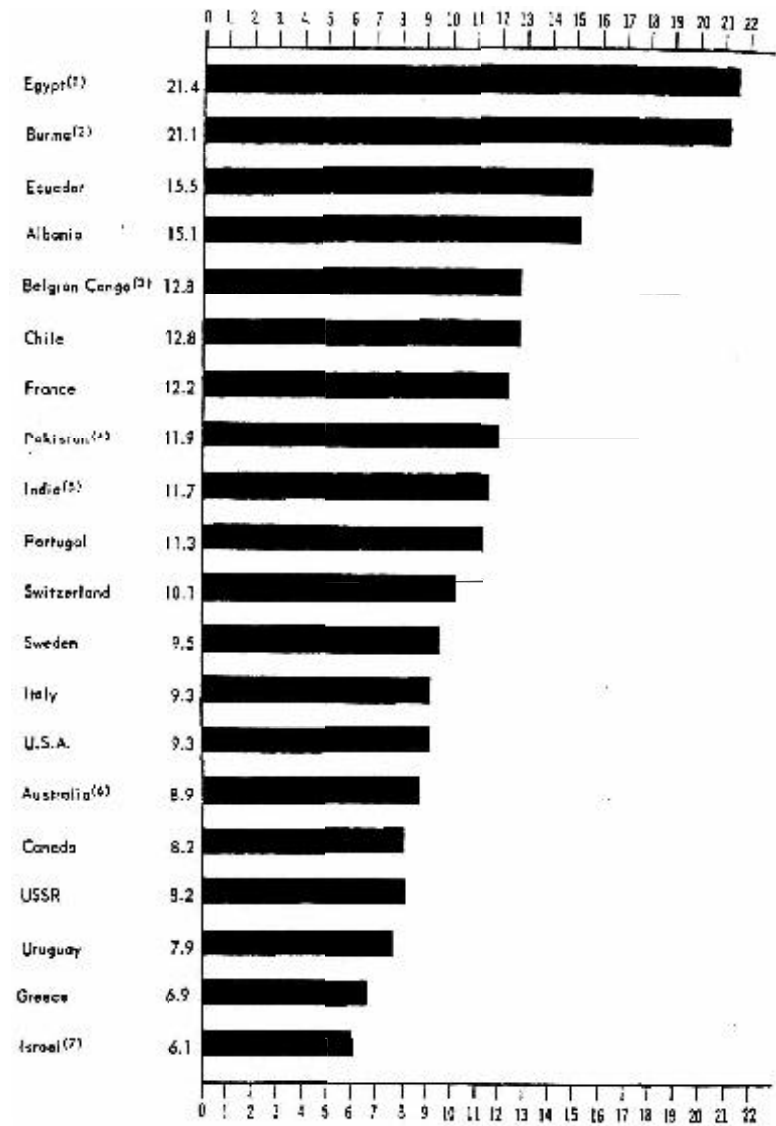
(1) 1954, Health Bureau; (2) 61 towns only; (3) Indigenous population; (4) 1951, Registration area, excluding live-born infants dying before registration of birth; (5) Registration area, excluding data for Bihar, Madhya Pradesh and Orissa; (6) Excluding full-blooded abirugubes; (7) Excluding Bedouin population in the Negev.

Source: Swaroop, S. Introduction to health statistics. E. & S. Livingstone Ltd., London, 1960.

2. Pictogram

The pictogram is a variation on the bar chart using a series of small identifying symbols to present the data. The figures are usually arranged horizontally, but may be arranged vertically. Each symbol may represent a single person or a particular unit. Generally, a symbol will represent a fixed number of persons or units. The number of items is determined by the variable being graphed, which in turn determines the length of the bar. Figure 15 illustrates a pictogram with a single person being represented by a symbol.

FIGURE 15: VACCINATION STATUS OF SMALLPOX CASES
UNITED KINGDOM - 1962, AND SWEDEN - 1963

YEARS SINCE
VACCINATION

NUMBER OF CASES
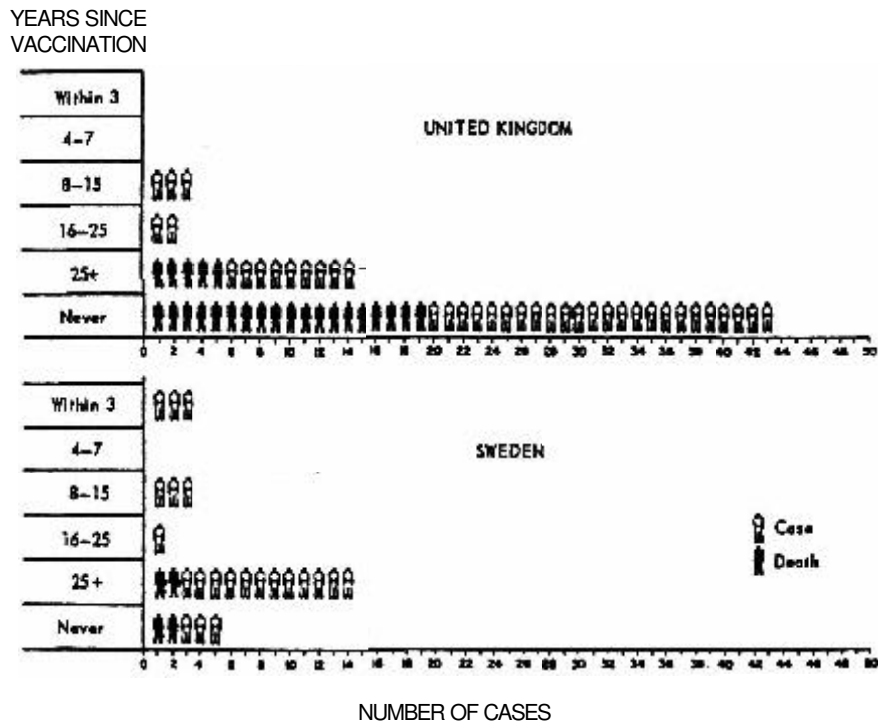
Source: SEP, NCDC

B. *Charts based on proportion*

Two types of the most important charts under this category are component bar charts and pie charts.

(1)   Component bar chart

A component bar chart is a bar chart in which the bars are divided into portions which are either coloured or shaded to denote their classifications. An example of a component bar chart is given in the following Figure 16.

**FIGURE 16: POLIOMYELITIS IMMUNIZATION STATUS
FOR CENTRAL CITIES (POP. ≥ 250 000)
BY AGE AND FINANCIAL STATUS
UNITED STATES, 1969**
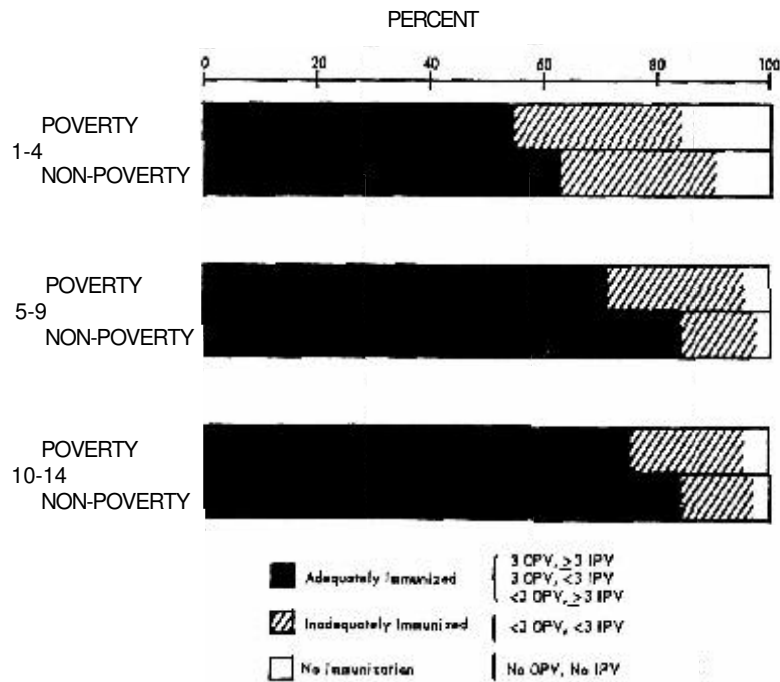
Source: 1969 U.S. Immunization Survey

(2)   Pie chart

Pie charts are charts which use wedge-shaped portions of a circle for comparison. This chart is sometimes called a sector chart. The pie chart is best adapted for illustrating the division of the whole into segments. The convention is to start at the 12 o'clock position and arrange segments in the order of their magnitude, largest first, and proceed clockwise around the chart. To convert from percentage to degrees, multiply the percentage by 3.6, since 360/100% = 3.6%. Figure 17 is an example of a pie chart.

**FIGURE 17: POLIOMYELITIS IMMUNIZATION STATUS OF CHILDREN
AGE 1-4 IN CENTRAL CITIES (POP. $\geq$250 000)
BY FINANCIAL STATUS
UNITED STATES, 1969**
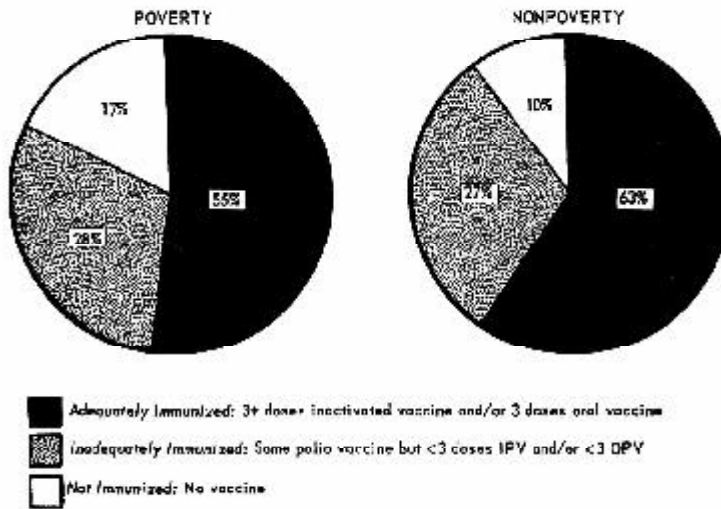
Source: 1969 U.S. Immunization Survey

C.   Geographic coordinate charts

Geographic coordinate charts are those which are based on geographical representation using maps. Disease outbreaks can be very easily plotted on a map for geographical spread. Also, percentages showing immunity levels, etc. can be shown in this fashion. Figure 18 is an illustration of a geographic coordinate chart.

**FIGURE 18: COUNTIES REPORTING ONE OR MORE CASES
OF ANIMAL RABIES
UNITED STATES, 1968**
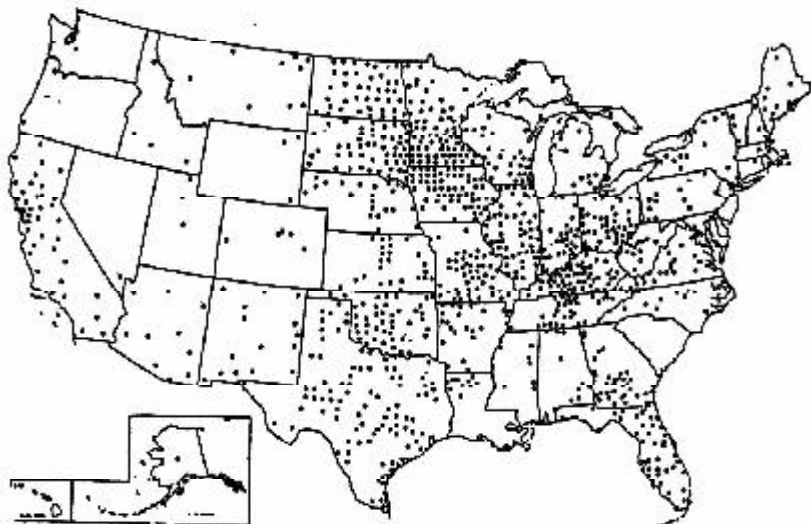
• COUNTY WITH ONE OR MORE CASES
   OF ANIMAL RABIES

Source: MMWR Annual Supplement, 1968.

D.   Special purpose charts

In addition to charts based on length, proportion and geographical coordinates, other visual methods of presenting data include flow charts, organizational charts, personnel charts and function charts.

1.   Flow charts

Examples of flow through a sewage treatment plant, flow through a water treatment plant, or such are examples of flow charts. Food flow through a restaurant could be presented visually in a flow chart as in Figure 19.

FIGURE 19: FOOD FLOW IN JOE'S RESTAURANT, SAMPLE CITY, 1970
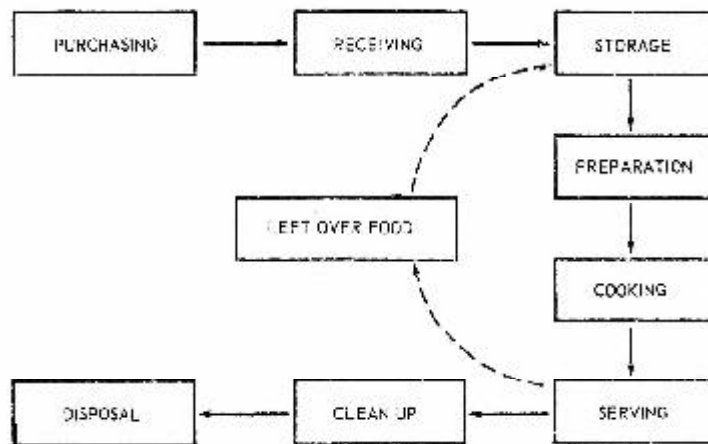
2. Organizational charts

Organization charts are sometimes confused with personnel charts and function charts. Organization charts should show the name of the office, division or section - not the personnel involved or their function. An example of an organization chart is shown is Figure 20.

FIGURE 20: ORGANIZATION CHART OF THE TILLER COUNTY HEALTH DEPARTMENT, 1968
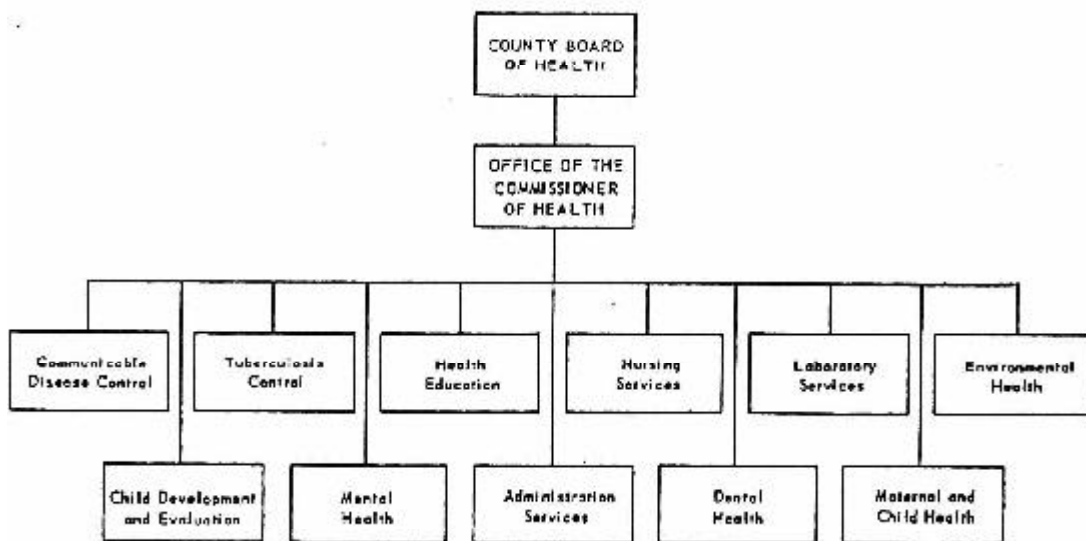
IV. Suggestions for the design and use of tables, graphs and charts

In conclusion, it would be well to review and reinforce what has been discussed up to this point.

A. Choose the tool most effective for data and purpose

Some methods of presentation call for more complete data than others; a few require special configurations of data. Within such limitations, decide upon the precise idea that you wish to communicate, then choose the method: continuous line graphs are suitable for a comparison of trends; bar charts clearly compare separate quantities of limited number; pie charts have advantages in comparing parts to their whole; scatter diagrams are excellent for showing tendency.

B.   Point out one idea at a time

Confine the presentation to one purpose or idea; limit the amount of data and include only one kind of data in each presentation. Different viewpoints on the information (unless they are being compared) call for separate presentations. So do large quantities of information or various kinds of information.

C.   Use black and white for exhibits that are to be reproduced

Few copying machines can reproduce colour -- and all colour reproduction is expensive. Colour can be adequately replaced: for areas, by cross-hatching or dotted fields; for lines, by continuous marks, dots, dashes (of different lengths in different lines), or combinations of the foregoing.

D.   Use adequate, properly located labels

Titles should include the 'what, where and when' that completely identify the data they introduce. All other labels should be as clear, complete and easy to understand -- and like the title, they should be outside the frame of the data. Only keys or legends (and these in a neat 'box' that sets them apart from the data) should appear within the field of a graph or chart.

E.   Give your sources

Where, or how (or both) the data were obtained is vital. Verification or further analysis by members of your audience is difficult if not impossible without full disclosure of sources. Also, access to the original information can prove as useful to the audience as either the excerpts that you present or the conclusions that you propose from them.

F.   Use care in proposing conclusions

In particular, draw conclusions that reflect the full body of information from which excerpted data were taken; propose only such conclusions as the data that you present can support. But keep in mind that tables, graphs and charts emphasize generalities - at the expense, of course, of detail. Compensate for this distortion, both in design and in comment. Footnote, in a prominent way, any important detail that has been obscured. Avoid conclusions that do not take such distortion into account.

References and further reading

Ipsen J., Feigl P. Bancrofft's introduction to biostatistics. New York, Harper and Row, 1970.

Schor S.S. Fundamentals of biostatistics. New York, G.P. Putnam's Sons, 1968.

Swaroop S. Introduction to health statistics. Edinburgh and London, E. & S. Livingstone Ltd., 1960.

Annex 3

Organization of a Workshop on Research Methods in Health Sciences

Background

During the Fourth Session of the Western Pacific Advisory Committee on Health Research (WPACHR), which was held in April 1979, a recommendation was made that the WHO Regional Office for the Western Pacific arrange workshops in research design and methodology in its member countries, with emphasis on the preparation of research grant proposals. Accordingly, 14 national workshops were held between 1981 and 1991 in collaboration with their national counterparts: four in China; two each in Malaysia, Papua New Guinea and the Philippines; and one each in Brunei Darussalam, Fiji, Republic of Korea and Viet Nam.

Doubts had previously been expressed by many people on whether research could be taught through short training courses, such as the national workshops organized by WHO. Such criticisms are fully understandable, since research demands a great deal of personal motivation, a wide range of knowledge and skills as well as creative ability, which can hardly be acquired by passive learning during short courses; yet, many people believe that junior scientists and medical practitioners can benefit from short courses focused on research design. If persons who have followed the courses can understand and recall the steps to be taken when they construct research projects, the gain may be great.

Many developing countries have recognized their research needs and are striving to develop their research capabilities. It is highly unlikely, however, that many of these countries will be able to acquire enough formally qualified, native epidemiologists and statisticians over the next 15 years to staff all the research programmes in need of their specified expertise. One may recall that such research efforts were a part of our attempt to achieve the goal of 'health for all by the year 2000'.

So what is a policy-maker to do? One approach is to organize the short courses for training in research methodology and at the same time increase the number of research specialists trained through long-term academic programmes. WHO has championed this approach, and more than 50 short courses for research methodology training were sponsored by the regional offices during the last decade, the largest share being that of SEARO, followed by WPRO. Both regions have elaborate organizational structures in the field of research planning and implementation.

Objectives

The major objective of such courses is to provide participants with a systematic approach to conducting research, in the expectation that they will disseminate this approach to the research community in general and transmit the knowledge and skills acquired during the workshop to junior researchers and trainees. The participants are expected to be able to carry out short training courses on research methodology at their own institutions, in order to transmit the rules in scientific research that should be followed closely by investigators to answer relevant research questions adequately, objectively and without bias.

After the course, the participants should therefore be able to:

- understand and appreciate better the scientific method as it is applied to research essential for proper clinical practice, prevention and control of diseases and health care delivery;

- give a precise statement about a problem to be investigated tand about the objectives of research into the problem;

- recognize the relevance of a problem to be investigated to national or local health development;

- frame pertinent hypotheses, which can then be tested by scientific methods to produce valid and useful results;

- construct a research proposal by selecting and applying the appropriate research design and methods;

- execute the research;

- properly interpret and effectively present the results to fellow scientists and investigators, policy-makers and administrators, and the public; and

- train junior scientists in the above concepts.

The framework for research methodology, which course participants are expected to have mastered, is shown in Table 1.

Working approach

The approach of the workshop should be to provide a broad framework for research methodology and design for use in biomedical research and their application in health systems. Usually, a topic is introduced by a lecture dealing with concepts and principles, followed by a practical workshop session and/or a presentation by the participants.

Duration of the workshop

The workshops last for two weeks.

Components of the workshop

1. Introduction to research:

   a. WHO research policy and national research coordination;

   b. introduction to research and scientific methods; definition, categories of research, scientific foundations of research, study design, planning and management of research; and

   c. concept of health systems research.

TABLE 1: RESEARCH METHODOLOGY FRAMEWORK

Research and scientific methods	• definition, role and scope of research • principles of science • inference and hypotheses • scientific proof and probability
Study design Surveys Analytical studies Trials Experiments	• study population • sampling • selection of controls • avoidance of bias • controlling of confounding factors
Constructing the research proposal	• defining the problem • framing the questions • stating the objectives • selecting the study design • planning the approach
Executing the study	• data collection • data collation • data processing • data analysis • interpretation of results
Presenting the results	• scientific publication • presentation at meetings • presentation for administrators and policy-makers • presentation to the public

2. A common system of research design and relevance to the development of research proposals using the WHO format

a. plan of study

- selection and formulation of research problems;

- appraisal of existing information;

- statement of objectives and research hypotheses; and

- research design and methodology for testing hypotheses.

b.   implementation of study

- collection of data;

- processing and analysis of data;

- interpretation and conclusions; and

- final report (presentation) and publication.

3. Selection of appropriate study design and research strategies

a.   descriptive strategies;

b.   analytical strategies: cohort studies and case-control studies; and

c.   experimental strategies: clinical trials and intervention studies.

4. Biostatistical support

a.   fundamental statistics;

b.   sampling and sample size;

c.   significance testing: hypothesis testing in statistics; and

d.   life-table techniques.

5. Practicums and preparation of proposals

a.   WPRO format for research proposals; and

b.   formulation of proposals (in groups).

6. Ancillary considerations

a.   bias and confounding;

b.   ethics in biomedical research; and

c.   data management and computers (field trip).

A flow chart of activities in a workshop on health research methodology and sample timetable are shown in Figures 1 and 2.

Participants

The number of participants should be around 25. These persons should hold relatively senior positions as:

1.   teaching staff in medical education, or

2.   a senior physician with genuine interest in developing research skills, or

3.   a research scientist with principal activities in the health field.

Faculty

The faculty should consist of experts in various disciplines appropriate to health research methodology.

A suggested module for course content

There is no fixed format for putting together the content of a short course in such a way that the course objectives will be satisfactorily fulfilled while the guidelines for a good course are followed. The module suggested here covers only possible content areas, which can be expanded or reduced as necessary according to course objectives, course duration and the background of the participants. Neither is there a fixed format for arranging the course material into sessions, day after day. This is, of course, a matter of style, common sense and logistics.

The module comprises three major instruction strata, namely: sessions on methodology, sessions for practicums and exercises, and sessions on substantive areas relative to the particular workshop.

1. Methodology component

a. Research design system

The core of the methodology component is the structure or system of research design, as outlined in Figure 3. The first four steps constitute the research proposal, while the subsequent four steps represent the conduct of the study, analysis and interpretation of data,

and preparation of scientific and progress reports. Each of these steps should be spelled out. More details should be given in the first four steps, and particularly in step 4. In the sessions devoted to describing the system, special items in step 4 (research plans) and step 7 (analysis of data) may be outlined sufficiently to be understood, while the details of these items may be left to the sessions on epidemiology, statistics or social science.

The research design system may be illustrated from literature relevant to the course participants, as time allows.

b. Epidemiological concepts and methods

Instruction in these areas should fit in and be relevant to the research design, depending on the course objectives, duration and participants' backgrounds. One must remember that this is not a course in epidemiology and, as such, is not intended or expected to produce 'instant epidemiologists'. The subject areas in epidemiology that are considered suitable for most short courses are:

•   measurement and epidemiological description;

•   various study designs (descriptive, analytical, clinical trials, experimental and evaluative);

•   assessment of risk and measures of effect for each study design; and

•   bias, confounding and causal inference.

The amount of detail given for these four areas will depend on the duration of the course. Additional concepts may be added in special courses.

c. Biostatistics

Instruction in biostatistics should follow the same general guidelines as for epidemiology. Short courses cannot afford a detailed treatment of many statistical concepts and procedures. As for epidemiology, all statistical instruction should fit into and refer to the research design. The subject areas in statistics that are considered suitable for most short courses are:

•   sampling procedures and randomization;

•   sample size determination; and

•   significance testing.

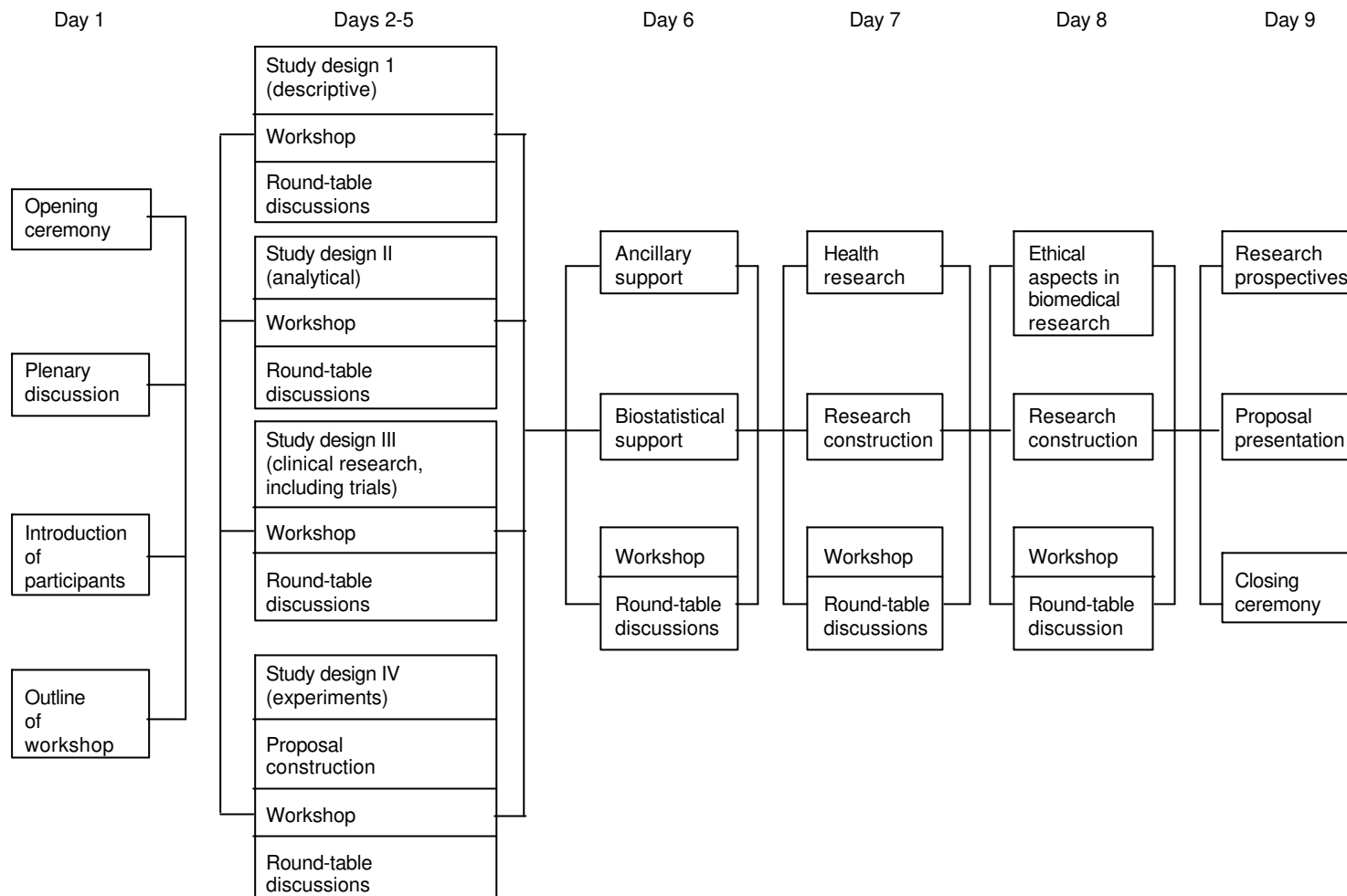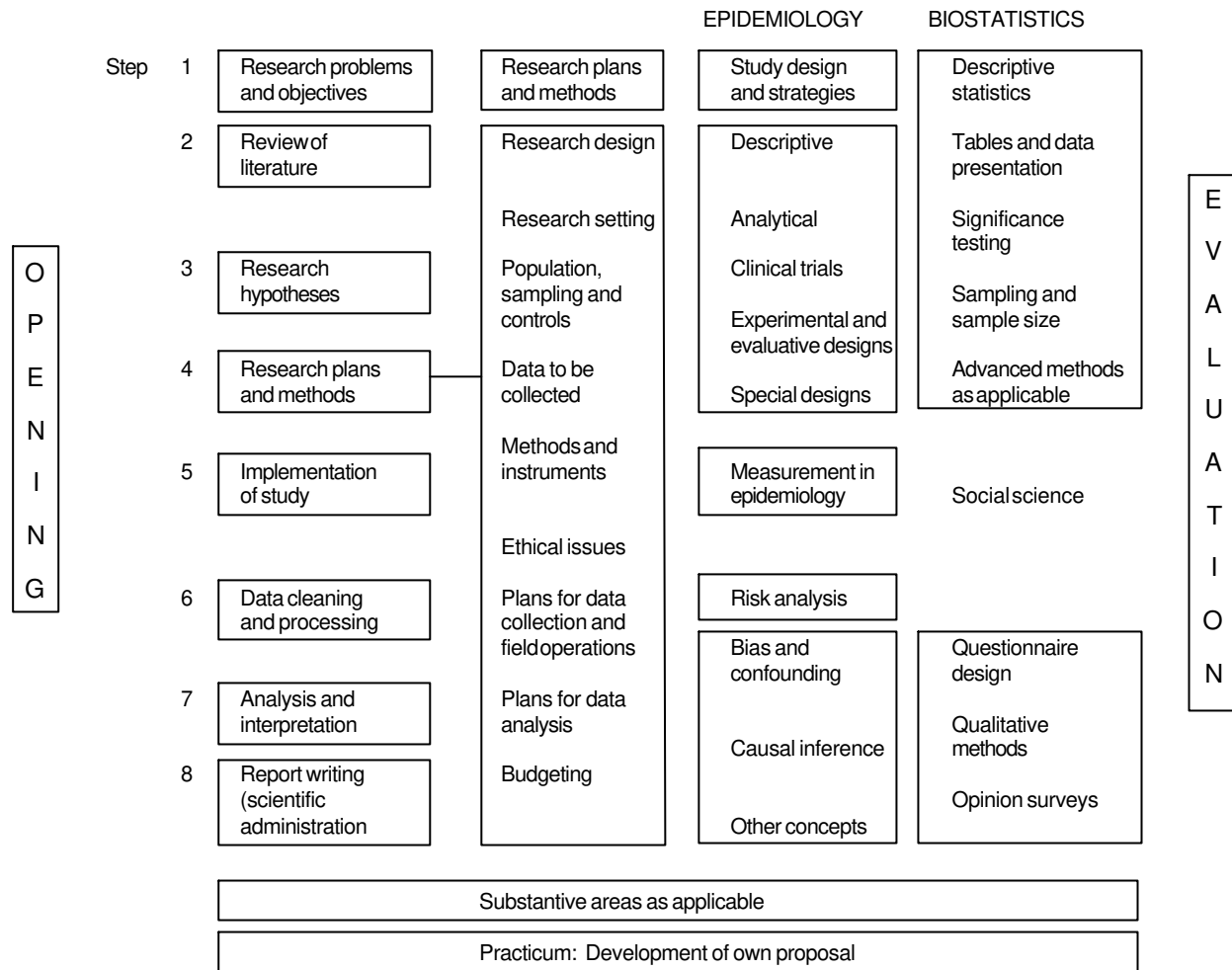**FIGURE 1: FLOW CHART OF ACTIVITIES AT A NATIONAL WORKSHOP ON HEALTH RESEARCH METHODOLOGY**

Day 1	Days 2-5	Day 6	Day 7	Day 8	Day 9
	Study design 1 (descriptive)				
	Workshop				
	Round-table discussions				
Opening ceremony	Study design II (analytical)	Ancillary support	Health research	Ethical aspects in biomedical research	Research prospectives
	Workshop				
	Round-table discussions				
Plenary discussion	Study design III (clinical research, including trials)	Biostatistical support	Research construction	Research construction	Proposal presentation
	Workshop				
Introduction of participants	Round-table discussions	Workshop	Workshop	Workshop	Closing ceremony
		Round-table discussions	Round-table discussions	Round-table discussion	
	Study design IV (experiments)				
Outline of workshop	Proposal construction				
	Workshop				
	Round-table discussions				

FIGURE 2: SAMPLE TIMETABLE - NATIONAL WORKSHOP ON RESEARCH DESIGN AND METHODOLOGY ON HEALTH RESEARCH, 18-30 AUGUST 1986

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6

8:30
 	
10:00
	Mon. (Aug. 18)
	
	Opening
	Framework
		Tues (Aug. 19)
		
		Research design
		system
		
		Step I
		       II
		       III
			Wed (Aug. 20)
			
			Research design
			
			Step V
			      VI
			      VII
			      VIII
				Thurs (Aug. 21)
				
				Diarrhoea study
				in the Philippines
					Fri (Aug. 22)
					
					Significance
					testing
						Sat (Aug 23)
						
						Reading on
						selected
						projects
						
						Discussion
						with faculty

10:15 11:45	WHO research programme Concept of health services research	Step IV	Review of descriptive statistics	Cohort studies Prospective Historical Prognostic	Questionnaire design and interview surveys	
1:15 2:30	Research methodology training	WHO format for research grant proposals	Life-table approach	How to present research results (tables, graphs)	Acute respiratory disease study in the Philippines	

2:45
 	
4:00
	Epidemiological
	description
	Disease
	   occurence
	Incidence/
	   prevalence
	Health indicators
	Introduction to
	   interpretation
		Illustrating
		research design
		by a clinical
		trial
			Introduction to
			strategies
			
			Descriptive
			Cross-sectional
			community
			   surveys
				Ethical issues
				
				Selection of
				research projects
				for participants
					Practicum
					(in groups)
					
					Problem
					identification
					Literature
					Hypotheses	

FIGURE 2: (CONTINUED)

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
	Mon. (Aug. 25)	Tues. (Aug. 26)	Wed. (Aug. 27)	Thurs. (Aug. 28)	Fri. (Aug. 29)	Sat. (Aug. 30)
8:30 10:00	Case-control studies	Regression, correlation (and orientation)	Clinical and field trials	Paper critique II in WHO and III	Visiting computer centre Group 1 orientation to the health information system	Presentation Group 2
10:15 11:45	Significance testing II	Bias and causal inference	Sample size determination	Writing scientific reports	(Faculty)	Closure
1:15 2:30	Sampling and controls	Paper critique I	Evaluation in health service research	Risk factors in atherosclerosis in Korean men	Practicum Plans for analysis (Dummy tables and graphs)	
2:45 4:00	Practicum Step I II III	Praticum Step IV	Practicum Samples/ Questionnaire	Practicum Questionnaire	Practicum Finalizing proposal (WHO format)	

FIGURE 3: SYSTEMS OF RESEARCH DESIGN

				EPIDEMIOLOGY	BIOSTATISTICS	
Step	1	Research problems and objectives	Research plans and methods	Study design and strategies	Descriptive statistics	
	2	Review of literature	Research design	Descriptive	Tables and data presentation	
			Research setting	Analytical	Significance testing	
	3	Research hypotheses	Population, sampling and controls	Clinical trials	Sampling and sample size	
	4	Research plans and methods	Data to be collected	Experimental and evaluative designs	Advanced methods as applicable	
				Special designs		
	5	Implementation of study	Methods and instruments	Measurement in epidemiology	Social science	
			Ethical issues			
	6	Data cleaning and processing	Plans for data collection and field operations	Risk analysis		
				Bias and confounding	Questionnaire design	
	7	Analysis and interpretation	Plans for data analysis	Causal inference	Qualitative methods	
	8	Report writing (scientific administration	Budgeting	Other concepts	Opinion surveys	

OPENING

EVALUATION

Substantive areas as applicable
Practicum: Development of own proposal

The amount of detail given will depend on the course objectives, its duration and the backgrounds of the participants. For beginners, we may add:

- descriptive statistics and data presentation; and

- orientation for special analytical techniques, including correlation, regression, multivariate analysis and life-table techniques. Few short courses can afford time for going into detail about these methods; exceptionally, one of the analytical procedures may be given in some detail. For example, in a course on studies of contraceptives, life-table techniques may be given in a special session to calculate effectiveness and continuation rate.

Experienced statisticians, social scientists or statistically trained epidemiologists may provide the statistical portion of a course, with examples and exercises from subject areas relevant to the audience. Instructors should be sensitive to the absorptive capacity of the audience and should tailor instruction to their level.

> Note: For unsophisticated audiences, the use of formulae may appear threatening and off-putting. They should be used sparingly. In sample size determination, reference to tables of sample size estimation may be made in addition to, or in lieu of, the various formulae. Participants should, however, know what information and assumptions are needed for determining sample size. They should be encouraged to seek statistical help during the development of their design, and not only after data collection.

d. Social science material

Instruction on social science material, such as qualitative studies, opinion surveys, questionnaire design and interviewing, can be given by faculty with experience in these areas. Details will again depend on course objectives, duration and participants.

2. Practicums and exercises (see Appendix)

Participants may be divided into small groups to work on practicums and exercises. Most crucially, participants should, individually or in small groups, work on developing a research proposal in an area relevant to their interests. Preceptors and resource persons are usually assigned to different groups. The proposal may, as a learning exercise, follow the steps of research design and should be begun as

early in the course as possible. The end product should be a written proposal. This exercise constitutes learning by doing, and it has been found that many methodological and conceptual issues become clearer during this session.

3. Substantive areas

Subject areas for which research designs are to be developed are known as substantive areas. A ready example is a course on research methods for clinical trials of contraceptive methods. The substantive area is contraception and fertility regulation from the physiological, clinical, cultural and service points of view. Again, adequate instruction in the substantive area is given so that a relevant research design can be developed.

4. Report writing

Although this area is an integral part of the system (step 8), it is worth a special session in short courses. Most of the candidates in these courses would appreciate guidelines for preparing not only scientific reports for publications, but also progress and final reports for administrative purposes.

Attributes of a 'good' curriculum for a short course

Short courses are demanding exercises for the planners, the faculty and the participants. Plans for the curricula of these courses vary according to local situations; however, several attributes should be considered in planning a curriculum.

1. The curriculum should be conceptually and functionally interdisciplinary, with a 'prescribed' share for each discipline according to the course objectives and dynamics.

2. The faculty should be collectively responsible for the course. Complementarity of instruction and smoothness of the flow of instruction from session to session are crucial. Flexibility may, nevertheless, be required, within reason.

3. The focus of the curriculum of a course for research methodology training should be on the process of research design. Epidemiological, statistical and other presentations should fit neatly within the research design system rather than being given as independent modules. This is elaborated on below. Certain courses may be -- by design - focused on specific

disciplinary areas, like courses on evaluative research, use of computers, epidemiological interpretation and health systems research.

4. The curriculum and material should be kept within the level of research experience and quantitative competence of the candidates. It should never be above their heads, nor should the details be beyond their absorptive capacity . The faculty should be encouraged to tailor their presentations and subject material to the particulars, duration and objectives of each course.

5. Educational objectives should be spelled out for each and every session of the course. Material to be presented, or outlines thereof, should be examined by the planning group to ensure that it will fulfil the educational objectives within the specified period. This approach can   be   a   most   influential mechanism for coordinating the course material and for eliminating unnecessary duplication or disciplinary bias. New faculty may consult the WHO   manual on instructional material for the procedures of preparing educational objectives.

6. Provisions should be made for small group learning, learning by doing, and frequent interaction among the participants on the one hand, and between the participants and faculty on the other. Further, the style of presentation should be stimulating and facilitating, as well as entertaining: dry, didactic lectures with no discussion have no place in short courses.

7. Recommendations and guidelines for further   learning resources and procedures may be given to ensure continuing education. Refresher courses for those with serious research interests may also be considered.

8. Projects selected for discussion and/or research consideration should be within the range of interest of the individuals or small groups. As far as feasible, the projects should be service-oriented.

9. The faculty should be recruited with great care. Teaching ability does not necessarily match scientific degrees. A rule of thumb is that the shorter the course and the less experienced the candidates, the greater the need for senior or experienced teachers.

10. Candidates should be selected according to specific criteria related to the objectives of the course. Haphazard selection and preference of favourites are counterproductive.

11. Attractive physical, social and logistic arrangements contribute to the success of short courses.

Flaws in short courses

The success of short courses in research methodology training is contingent upon their proper planning and conduct; however, not all short courses are as successful as one would like them to be. Several flaws that exist render the whole approach vulnerable to criticism; recognizing these flaws is the first step in an effort to improve the performance of the courses, if preventive and corrective measures are taken. The flaws include:

1. inflated objectives without realistic appreciation of the limitations inherent in short courses;

2. a disciplinary bias, whereby some planners may push for more representation of their own disciplines rather than coordinating their contribution with that of persons from other disciplines to achieve the course objectives;

3. inadequate treatment or glossing over of the process of research design, which should be central to research methodology and training;

4. monotonous use of the same subject material (and even the same hand-outs) with little modification of the course objectives, course duration or candidates' backgrounds;

5. giving presentations that are too complex for the participants;

6. recruitment of faculty who may have the scientific qualifications but lack teaching ability, especially for the demanding dynamics of short courses;

7. poor or haphazard recruitment of candidates; and

8. finally, a serious flaw that requires longer discussion. This flaw relates to conflicting, uncoordinated use of disciplinary jargon. There is nothing wrong with using jargon. On the contrary, faculty are encouraged to make disciplinary jargon familiar to participants, provided that it is relevant to research design. The problem arises with terms that are used differently in different disciplines. If participants are not cued to the existence

of these differences, confusion and frustration may result. Two frequent examples are 'hypothesis' and 'hypothesis testing', which are used differently by statisticians and epidemiologists.

Conclusions

1. Short courses in research methodology training are both a viable and feasible mechanism for strengthening research capability in many countries. They should not, however, preclude or be considered an exclusive alternative to long-term training.

2. The central core of a course should be research design. Epidemiology, biostatistics and social science are only tools that should be integrated into the research design.

3. Because of its brevity, this kind of course should be most carefully planned to optimize contributions and achieve course objectives.

4. The course should be tailored to the audience, which should be carefully selected.

5. Faculty should also be carefully selected and should be collectively responsible for the course.

Appendix

Assignment of articles for practicums and exercises

One or two articles should be selected from each of the following groups of papers suggested for cross-sectional, cohort and case-control studies.

1. Cross-sectional studies

Cristina Leske, M. (1981) Estimating incidence from age-specific prevalence in glaucoma. Am. J. Epidemiol., 113

Hallgrimsson, J. (1979) Chronic non-rheumatic aortic valvular disease: a population study based on autopsies. J. chron. Dis.,32, 355-363

Marmot, M. (1975) Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: prevalence of coronary and hypertensive heart disease and associated risk factors. Am. J. Epidemiol., 102

2. Cohort studies

Herbst, A. (1980) A comparison of pregnancy experience in DES-exposed and DES-unexposed daughters. *J. Reprod. Med.* 24, 62-69

Patriarca, P. (1982) Kawasaki syndrome; association with the application of rug shampoo. *Lancet*, ii

Seyfried, P. L. (1985) A prospective study of swimming-related illness. 1. Swimming associated health risk. *Am. J. Public Health*, 75, 1068-1070

The Coronary Drug Project Research Group (1979) Cigarette smoking as a risk factor in men with a prior history of myocardial infarction. *J. Chronic Dis.*, 32, 415-425

Vessey, M. P. (1977) Mortality among women participating in the Oxford Family Planning Association contraceptive study. *Lancet*, ii

Wahdan, M. H. (1980) A controlled field trial of live oral typhoid vaccine Tyzla. Bull. World Health Organ., 53, 469-474

Warrell, D. A. Dexamethasone proves deleterious in cerebral malaria. *New Engl. J. Med.*, 6, 205-211

Zumrawi, F. (1981) Dried skimmed milk, breast-feeding and illness episodes - a controlled trial in young children in Khartoum province, Sudan. *Int. J. Epidemiol.*, 10, 303-308

3. Case-control studies

Aromaa, A. (1976) Breast cancer and use of rauwolfia and other antihypertensive agents in hypertensive patients: a nation-wide case-control study in Finland. *Int. J. Cancer*, 18, 727-738

Beattie, A. D. (1975) Role of chronic low-level lead exposure in the etiology of mental retardation. *Lancet*, i, 589-592

Greenberg, G. (1977) Maternal drug histories and congenital abnormalities. *Br. Med. J.*, ii, 853-856

Halsey, N. A. (1980) Risk factors in subacute sclerosing panencephalitis. A case-control study. *Am. J. Epidemiol.*, iii, 415-424

Heinonen, O. P. (1974) Reserpine use in relation to breast cancer. *Lancet*, ii, 675-677

Nelson, D. B. (1980) Aflatoxin and Reyes' syndrome: a case-control study. *Pediatrics*, 66, 865-869

Shapiro, C. (1985) A case-control study of BCG and childhood tuberculosis in Cali, Colombia. *Int. J. Epidemiol.*, 14, 441-446

Shapiro, S. (1980) Recent and past use of conjugated estrogen in relation to adenocarcinoma of the endometrium. *New Engl. J. Med.*, 303, 485-489

Exercise format

Read the assigned article carefully and then use the following set of questions in your critique. If you miss questions, or if some of the concepts are not familiar to you, you should not worry. You will get these concepts during the course. This is only a probe. Please do not consult with anyone in your response.

1. Problem

Was the research problem clearly identified and stated?

____ Yes   ____ No

Were the specific questions for research specified, or can they be implied?

____ Yes   ____ No   ____ Implied

State the research problem in one paragraph.

_____

Formulate the two important specific research questions investigated in this study.

a.

b.

2. Objectives

Were the immediate and ultimate objectives of the study specified?

____ Yes   ____ No

Comments _____

3. Literature

Was the literature updated?

____ Yes   ____ No

Comments _____

Was the literature critically appraised?

____ Yes   ____ No

Comments _____

4. Hypotheses

Were the hypotheses underlying the study specified?

____ Yes   ____ No

State the hypotheses relevant to this study, in your own words.

a.   The conceptual hypothesis:

b.   The operational hypothesis:

5. Strategy

How would you classify this study: (choose one)

a.   randomized controlled double-blind clinical trial

b.   randomized controlled clinical trial

c.   controlled clinical trial

d.   descriptive study, ecological study

e.   case-control study

f.   cohort study (prospective, historical)

g.   cross-sectional study

h.   other (specify)

6. Population

Which were the units of observation in this study?

7. Data to be collected

What are the data to be collected about these units?

Which are the independent and dependent variables?

8. Sampling

What was the sampling technique?

Were the subjects selected as representative samples?

To which population can the results of this study be generalized?

9. Control

What controls were used?

Were the study and control groups reasonably comparable?

____ Yes   ____ No

Comments _____

If you were to redesign this study, how would you allocate at random to the study and control groups?

What criteria were used to show comparability of the study and control groups?

10. Study instruments

What were the study instruments used in this study?

a. questionnaire

b. interview schedule

c. medical examination

d. laboratory procedures

e. other: specify

Were ethical problems adequately considered?

_____ Yes   _____ No

Comments _____

11. Flow chart

Draw a flow chart summarizing the study.

12. Schedule

What was the duration of the study?

13. Analysis

Were the methods of analysis adequate in regard to:

a. tabulation   _____ Yes   _____ No

Comments _____

b. significance testing   _____ Yes   _____ No

Comments _____

c. control of confounding   _____ Yes   _____ No

Comments _____

Considering the statistical data analysis, answer the following questions:

a. which statistical tests were used? Explain results.

b. what significance level was used?

c. what other statistical tests could be suggested?

14. Conclusions

Did the conclusions flow logically from the results of the study, or were they biased?

Are the results of this study consistent with what you know from personal research, experience or reading?

Would you accept the results of this study? (choose one)

a. yes - unquestionably

b. yes - with some reservations

c. no

d. comments:

15. Overall

What is your overall evaluation of this study in terms of strengths and limitations?

a.   Strengths:

b.   Limitations:

References and further reading

Guilbert J.J. Educational handbook for health personnel, revised ed. Geneva, World Health Organization, 1981 (WHO offset publication no. 35).

Omran A.R. The Clark-Omran system of research design in epidemiology. Raleigh, NC, University of North Carolina, 1972.

Paik Y.H., Patwary K.M. Report on the National Workshop on Biomedical Research Methodology, Chun-Chon City, Kwangdo, Republic of Korea, 20-28 October 1983. Manila, WHO Regional Office for the Western Pacific, 1983 (ICP/RPD/002).

Pauli H.G. Training in research methodology: review and proposals (Advisory Committee on Medical Research, 25th Session, Geneva, 10-13 October 1983). Geneva, World Health Organization, 1983.

Index