LECTURE NOTES

For Health Science Students

Research Methodology

Getu Degu

Funded under USAID CooperativeAgreement No. 663-A-00-00-0358-00.

Produced in collaboration with the Ethiopia Public Health Training Initiative, The Carter Center, the Ethiopia Ministry of Health, and the Ethiopia Ministry of Education.

Important Guidelines for Printing and Photocopying
Limited permission is granted free of charge to print or photocopy all pages of this publication for educational, not-for-profit use by health care workers, students or faculty. All copies must retain all author credits and copyright notices included in the original document. Under no circumstances is it permissible to sell or distribute on a commercial basis, or to claim authorship of, copies of material reproduced from this publication.

©2006 by Getu Degu and Tegbar Yigzaw

In general, this lecture note tries to cover the three major components of a research process: development of the research proposal, fieldwork (data collection) and write-up of the scientific report. General learning objectives followed by introductory sections which are specific to each chapter are placed at the beginning of most of the chapters. The lecture note also includes a number of exercises for the students so that they can examine themselves whether they have understood the topic under consideration. It is assumed that this lecture note on research methodology will be given to health science students who have taken basic Epidemiology and Biostatistics courses.

It is also important to note that this lecture note focuses on quantitative research. When the point of discussion refers to qualitative research, it would be clearly shown so as to avoid the confusion that may arise. A few reference materials are given at the end of the lecture note for further reading.

ACKNOWLEDGMENTS

We would like to thank the College of Medicine and Health Sciences (University of Gondar) for allowing us to use the resources of the institution while writing this lecture note. We are highly indebted to the Carter Center (Ethiopian Public Health Training Initiative) without whose support this material would have not been written. In particular, we are very grateful to Ato Aklilu Mulugeta (from the Carter center) for his uninterrupted follow up and encouragement. We would like to extend our gratitude and appreciation to Dr. Getnet Mitikie and Dr. Mesganaw Fantahun of Addis Ababa University Associate and Assistant Professors respectively for their critical reviews and valuable comments on the initial draft of these teaching materials. .

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION TO RESEARCH

1.1 Learning Objectives

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

1. Define research in general and health systems research in particular
2. Enumerate the characteristics of research
3. Identify the different types of research
4. List the essential features of health systems research
5. Describe the broad divisions (steps) involved in the research process
6. Explain the roles of research in development

1.2 Introduction

behave in a certain way but they do not adequately answer the “why” and “how” questions. Research which attempts to increase our understanding of why things are the way they are in our social world and why people act the ways they do is “qualitative” research.

Qualitative research is concerned with developing explanations of social phenomena. That is to say, it aims to help us to understand the world in which we live and why things are the way they are. It is concerned with the social aspects of our world and seeks to answer questions about:

- Why people behave the way they do
- How opinions and attitudes are formed
- How people are affected by the events that go on around them
- How and why cultures have developed in the way they have

Qualitative research is concerned with finding the answers to questions which begin with: why? How? In what way? Quantitative research, on the other hand, is more concerned with questions about: how much? How many? How often? To what extent? etc.

Public health problems are complex, not only because of their multicausality but also as a result of new and emerging domestic and international health problems. Social, economic, political, ethnic, environmental, and genetic factors all are associated with today’s public health concerns. Consequently, public health practitioners and researchers recognize the need for multiple approaches to understanding problems and developing effective interventions that address contemporary public health issues. Qualitative methods fill a gap in the public health toolbox; they help us understand behaviors, attitudes, perceptions, and culture in a way that quantitative methods alone cannot. For all these reasons, qualitative methods are getting renewed attention and gaining new respect in public health.

A thorough description of qualitative research is beyond the scope of this lecture note. Students interested to know more about qualitative methods could consult other books which are primarily written for that purpose. The main purpose of this lecture note is to give a detailed account on the principles of quantitative research.

Health research

Health research is the application of principles of research on health. It is the generation of new knowledge using scientific method to identify and deal with health problems. Knowledge, both generalizable worldwide and locally specific, is essential to effective action for health. Worldwide knowledge is the basis on which new tools, strategies, and approaches are devised that are applicable to health problems facing many countries. Local knowledge, specific to the particular circumstances of each country can inform decision regarding which health problems are important, what measures should be applied and how to obtain the greatest health benefit from existing tools and limited resources. In this regard, health research is both global and local in nature.

In most cases, health research has been divided into three overlapping groups.

Essential health research : Consists of activities to define the health problems of a given country or community, to measure their importance and to assure the quality of activities to deal with them. Much of this research comes within the category of health service research but there will be elements of clinical research and development of technology, depending on the situation. The information, which may be obtained in a number of ways, is essential and specific to each country for planning and monitoring health services. Some of the research conclusions, however, may be generalized and applicable to other areas.

Clinical research : In its widest sense, this group of topics ranges from studies of the prevention and diagnosis of diseases through new methods of treatment to problems of care and rehabilitation. The sophistication will vary from problem to problem and there will be overlap with the fields of essential and biomedical research. Some of the research will be mainly of local importance; much will be useful for other individuals in other countries. Examples include clinical trials of disease prevention and the design of new chemotherapeutic agents. Wherever clinical facilities exist, there is a potential for clinical research.

1. The individual, family and the community
2. Health care services private sector: traditional and modern medical practice (legal or illegal)

Public (governmental) sector

Health workers, health institutions, etc.

3. Health related sectors education, agriculture, etc.
4. The international sector, including bilateral and multilateral donor agencies (UNICEF, WHO, etc.) that may support health as well as Essential Features of Health Systems Research (HSR)

Bearing in mind that HSR is undertaken primarily to provide information to support decision-making at all levels that can improve the functioning of the health system, some of the essential features are summarized as follows:

- HSR should focus on priority problems .
- It should be action oriented (i.e., aimed at developing solutions)
- An integrated multidisciplinary approach is required (research approaches from many disciplines)
- The research should be participatory in nature (from policy makers to community members)
- Research must be timely.
- Emphasis should be placed on comparatively simple, short-term research designs

- Honest discussion of practical or methodological problems that could have affected the findings.
- Alternative courses of action that could follow from the results and the advantages and drawbacks of each.

9. Evaluation of the research undertaken - An HSR project should not stop at finding answers to the research questions posed, but include an assessment of what decisions have been made based on t

CHAPTER TWO
TOPIC SELECTION

2.1 Learning objectives

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

1. Examine the cyclical nature of the development of a research proposal
2. Describe the principles underlying whether a problem situation is researchable.
3. List the criteria for selecting a research topic.
4. Identify and select his/her own topic (health problem) for research based on certain guidelines.

2.2 Introduction

The development of a health project goes through a number of stages. Formulation of the research proposal is the major task in the process of developing a research project. The proposal draws on all the preparatory steps of the research process and pulls them together in a document describing the rationale and the methodology proposed for research. The proposal is a basis for approval and funding. After approval, the proposal is used as a blueprint during implementation of the project. It should be noted that development of a research proposal is often a cyclical process. The process is not always linear. It is a usual practice to go up and down on the developed proposal and make the necessary revisions.

Is there evidence to indicate that the research proposal focuses on a problem of priority importance? Was the given health problem identified by relevant groups of the health system? Was the problem adequately analysed to include all possible contributory factors from different sectors? Was it clearly stated? These questions should be clearly answered before trying to develop the research proposal. The sections that follow are devoted to giving the guidelines useful for identification, selection, analysis and statement of the given problem.

2.3 Problem identification

If the answer to the research question is obvious, we are dealing with a management problem that may be solved without further research. A number of research questions could be presented that may be posed at the various levels of the health system.

Whether a problem requi res research depends on three conditions:

I) There should be a perceived difference or discrepancy between what it is and what it should be;

II) The reason(s) for this difference should be unclear (so that it makes sense to develop a research question); and

III) There should be more than one possible and plausible answer to the question (or solution to the problem).

example1:

Problem situation: In district " Y " a report showed that in the first month there were 500 children under one year old who started immunization, but at the end of the year it was found out that there were only 25 children who completed their vaccination.

Discrepancy: All the 500 children at district "Y "

Example 2:

Problem situation: In district “Z” (population 150,000) there are 2 health centers, 1 hospital and 15 health stations and all of them function smoothly. However, at the end of the year it was found that the EPI coverage was only 25%.

Discrepancy: Although district “Z” had 100% availability of health services and at least 80% of the children should have had full vaccinations the EPI coverage was only 25% as seen above.

Problem question: What factors influence the low EPI coverage in district “Z”?

Possible answers:

- Mothers might have problems for not attending in the EPI sessions.
- The MCH, EPI, OPD, CDD, etc… programmes might not have been integrated; hence children might have missed opportunities in getting immunization.
- The follow up of defaulting children might not be effective and other reasons.

Thus, the above problem situation is researchable .

2.4 Criteria for prioritizing problems for research

Each problem that is proposed for research has to be judged according to certain guidelines or criteria. There may be several ideas to choose from.

Before deciding on a research topic, each proposed topic must be compared with all other options.

The selection and analysis of the problem for research should in volve those who are responsible for the health status of the community. This would include managers in the health services, heal th-care workers, and community leaders, as well as researchers.

The guidelines or criteria given below can help in the process of selection.

6. Urgency of data needed

How urgently are the results need ed for making a decision? Which research should be done first and which can be done late?

7. Ethical acceptability

We should always consider the possibilit

Urgency

Rating Sheet

Criteria for selecting a research topic	Proposed topic		
	Health problem I	Health problem II	Health problem III
Relevance			

a) Why is it important to stat e and define the problem well?

Because a clear statement of the problem:

- Is the foundation for the further development of the research proposal (research objectives, methodology, work plan, etc);
- Makes it easier to find information and reports of similar studies from which your own study design can benefit;
- Enables the researcher to systematically point out why the proposed research on the problem should be undertaken and what you hope to achieve with the study results.

b) Points that need to be considered for justifying the selected research problem

A health problem selected to be studied has to be justified in terms of its:

- Being a current and existing problem which needs solution
- Being a widely spread problem affecting a target population
- Effects on the health service programmes
- Being a problem which concerns the planners, policy makershich] .725 TD 0 T

- An analysis of the major factors that may influence the problem and a convincing argument that available knowledge is insufficient to answer a certain question and to update the previous knowledge.

- A brief description of any solutions that have been tried in the past, how well they have worked, and why further research is needed.

- A description of the type of information expected to result from the project and how this information will be used to help solve the problem

- If necessary, a short list of definitions of crucial concepts used in the statement of the problem.

A list of abbreviations may be annexed to the proposal, but each abbreviation also has to be written out in full when introduced in the text the first time.

3.5 Exercises

1. Why do we need to analyze the research problem?
2. What are the points required to justify the selected research problem?
3. What information should be included in the statement of a problem?

CHAPTER FOUR
LITERATURE REVIEW

4.1 Learning objectives

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

1. Describe the reasons for reviewing available literature and other information during the preparation of a research proposal.
2. Describe the resources that are available for carrying out such a review.
3. Record (organize) information obtained from literature on an index card.

4.2 Introduction

At the outset of his/her study the investigator should be acquainted with the relevant literature. It is of minimal use to wait until a report is written.

4.3 Use of literature review

- It prevents you from duplicating work that has been done before.
- It increases your knowledge on the problem you want to study and this may assist you in refining your "statement of the problem".
- It gives you confidence why your particular research project is needed.
- To be familiar with different research methods

4.4 Sources of information

- Card catalogues of books in libraries
- Organizations (institutions)
- Published information (books, journals, etc.)
- Unpublished documents (studies in related fields, reports, etc.)
- Computer based literature searches such as Medline
- Opinions, beliefs of key persons

Some examples of resources where in formation could be obtained are:

- Clinic and hospital based data from routine activity statistics
- Local surveys, annual reports
- Scientific conferences
- Statistics issued at region and district levels
- Articles from national and international journals (e.g., The Ethiopian Journal of Health Development, The Ethiopian Medical Journal, The East African Medical journal, The Lancet, etc.)
- Internet
- Documentation, reports, and raw data from the Ministry of Health, Central Statistical Offices, Nongovernmental organizations, etc.

References that are identified:

- Should first be skimmed or read
- Then summaries of the important information in each of the references may be recorded on separate index cards
- These should then be classified so that the information can easily be retrieved

4.5 Organization of information on index cards

The index cards should contain:

- Key words
- A summary of the contents of books or articles which is relevant to one's own study
- A brief analysis of the content, with comments such as:
  - how information from that particular study could be used in one's own study

xx

- from past to current

In conclusion, while reviewing a literature, all what is known about the study topic should be summarized with the relevant references. This review should answer

- How much is known?
- What is not known?
- What should be done based on what is lacking?

Overall, the literature review should be adequate, relevant and critical. In addition to this, appropriate referencing procedures should always be followed in research proposals as well as in research reports. While reviewing a literature give emphasis to both positive and negative findings and avoid any distortion of info

CHAPTER FIVE

OBJECTIVES

5.1 Learning objectives

After completi ng this chapter, the student should be able to :

1. Describe the need for the development of research objectives
2. Differentiate between general and specific objectives
3. Formulate specific objectives and hypotheses

5.2 Introduction

Having decided what to study, and knowing why s/he wants to study it, the investigator can now formulate his study objectives. Objectives should be closely related to the statement of the problem. For example, if the problem identified is low utilization of health stations in a rural district, the general objective of the study could be to assess the reasons for this low utilization. If we break down this general objective into smaller and logically connected parts, then we get specific objectives.

5.3 Definitions

General objectives : aim of the study in general terms

Example: In a study on missed opportunities for EPI in Addis Ababa the general objective was: “to assess missed opportunities fo r EPI in Addis Ababa”.

Specific objectives: measurable statements on the specific questions to be answered. Unlike the general objectives, the specific objectives are more specific and are related to the research problem situation. They indicate the variable to be examined and measured.

Example: In the study of missed opportunity for EPI in Addis Ababa the specific objectives could be:

- To find out the magnitude of missed opportunities for children who attend OPD, MCH, CDD, etc. in Addis Ababa,

- To examine the reasons for children not being immunized while attending the OPD, MCH, CDD, etc. services.

5.4 Formulation of the research objectives

The formulation of objectives will help us to:

- Focus the study (narrowing it down to essentials)
- Avoid collection of data that are not strictly necessary for understanding and solving the identified problem
- Organize the study in clearly defined parts

The explicit formulation of study objectives is an essential step in the planning of a study. It is said that “a question well-stated is a question half-answered”, but a question that is poorly stated or unstated is unlikely to be answered at all.

How should we state our objectives?

We have to make sure that our objectives:

- Cover the different aspects of the problem and its contributing factors in a coherent way and in a logical sequence
- Are clearly expressed in measurable terms
- Are realistic considering local conditions
- Meet the purpose of the study
- Use action verbs that are specific enough to be measured

Examples of action verbs are:

- to determine
- to compare
- to verify

- to calculate
- to describe
- to find out
- to establish

Avoid the use of vague non-action verbs such as;

- to appreciate
- to understand
- to study
- to believe

Research objectives can be stated as:

- Questions - the objectives of this study are to answer the following questions ….
- Positive sentence - the objectives of this study are to find out, to establish, to determine, …
- Hypothesis - the objective of this study is to

3. An increase in the frequency of face washing is followed by a reduction in trachoma prevalence

One of the most important problems usually observed among students is the tendency of stating too many study objectives which are not appropriately addressed (or sometimes will be forgotten) in the sections that follow. It should be noted that it is on the bases of these specific objectives that the methods, results and discussion sections will be presented. For example, sample size calculations for each stated objective and identifying (selecting) the most appropriate sample size that will answer the required research questions is not covered in the development of most research proposals. This is also true during the write up of the completed research work. It is not uncommon to come across a situation in which some of the specific objectives are not addressed in the results section at all. It is therefore advisable to limit the number of specific objectives. In most practical situations, the number of specific objectives should not exceed three.

5.5 Exercises

1. Define general objectives, specific objectives and hypotheses by giving your own examples.
2. The objectives of a study should be written after the statement of the research problem and before the methods section. Does this statement sound true? Justify your answer.
3. List the characteristics of research objectives.
4. Comment on the statement: “A question well-stated is a question half-answered”.
5. Mention some of the problems that may arise as a result of having too many objectives.

CHAPTER SIX
RESEARCH METHODS

6.1 Learning objectives

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

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

6.2 Introduction

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

6.3 Types of study designs

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

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

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

Uses of descriptive studies

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

Types of descriptive studies

a) Case reports and case series

Case report: a careful, detailed report by one or more clinicians of the

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

ƒ Ecological studies are usually quick and easy to do and can be done with already available information.

ƒ Since ecological studies refer to whole popula

studies provide the required data at more than one point in time unlike cross- sectional surveys).

II. Analytic studies

Analytic studies may be defined as studies used to test hypotheses concerning the relationship between a suspected risk factor and an outcome and to measure the magnitude

Strengths and limitations of the cohort study design

Strengths:

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

Limitations:

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

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

Strengths and limitations of the case-control study design

Strengths:

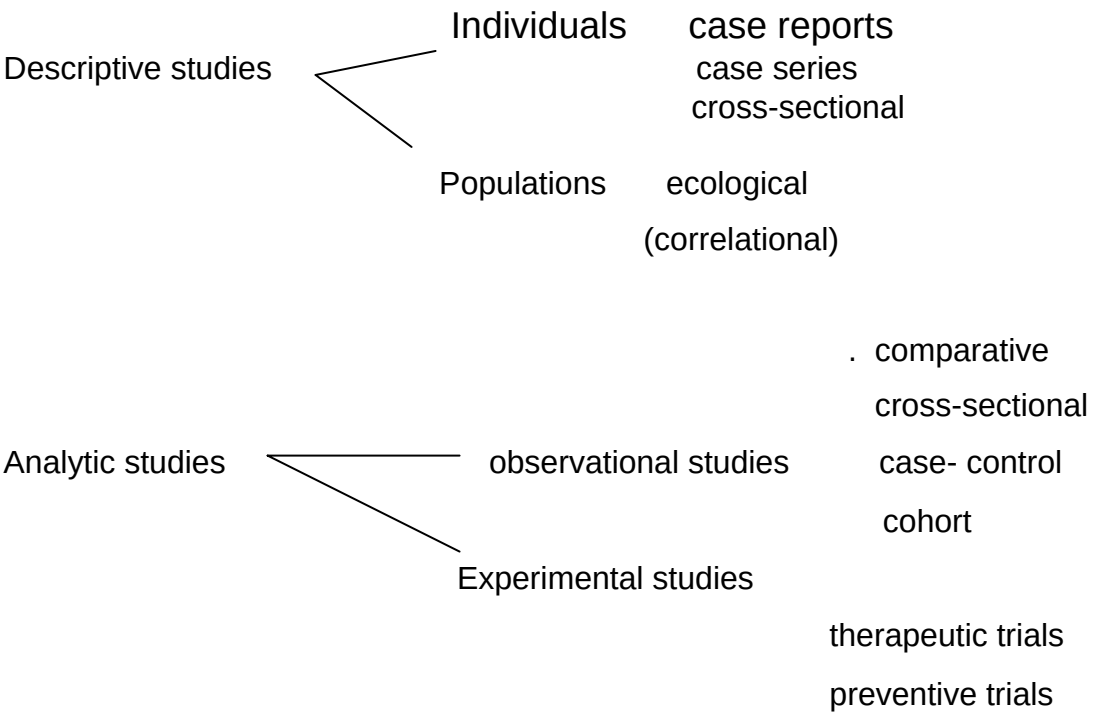
- Is relatively quick and inexpensive
- Is optimal for the evaluation of rare diseases.
- Can examine multiple etiologic factors for a single disease.

Limitations:

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

One of the most common quasi-experimental designs uses two (or more) groups, one of which serves as a control group in which no intervention takes place. Both groups are observed before as well as after the intervention, to test if the intervention has made any difference. (This quasi-experimental design is called the ‘non-equivalent control group design’

Summary



6.4 Study population

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

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

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

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

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

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

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

Practical questions such as th e following could also arise .

- Is the proposed population the one that would give the required information?
- Will the population cooperate to participate in the study, or will it be a 'resistant' one?
- If it is proposed to study patients with a specific disease, will it be possible to identify enough cases to yield useful conclusions?
- If a long term 'follow up' study is planned, is the population so mobile that it may be difficult to maintain contact with the subjects?

A preliminary exploratory study may sometim es be required in order to answer such questions.

6.5 Operational Definitions of Variables

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

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

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

a) weight in kilograms
b) height in centimeters

religion will be included. These background variables are often related to a number of independent variables, so that they influence the problem indirectly. Hence they are called background variables or background characteristics.

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

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

I. Operationalising variables by choosing appropriate indicators

Note that the different values of many of the variables presented above can easily be determined. However, for some variables it is sometimes not possible to find meaningful categories unless the variables are made operational with one or more precise INDICATORS. Operationalising variables means that you make them ‘measurable'.

For example:

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

— 0 to 3 correct answers is poor,

— 4 to 6 correct answers is reasonable, and

— 7 to 10 correct answers is good.

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

- The initial list is usually too long
- It has to be pruned to facilitate th e collection and processing of the data.

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

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

Scales of Measurement

As part of the process of clarifying each of the variables to be studied, its scale of measurement should be specified. There are four types of scales of measurement: Nominal, Ordinal, Interval and Ratio. They are listed in ascending order of power and preference.

1. Nominal Scale: This consists of two or more named categories (classes) which are qualitatively different from each other.
   E.g Sex: male (1); Female (2)
   Marital status: 1. Married 2. Single 3. Divorced 4. Widowed

2. Ordinal scale: This has the additional quality that the categories are ranked and have implied order. However, the intervals between classes are not necessarily equal.

   Example 1. Severity of a disease: Severe (grade III); moderate (grade II); mild (grade I); absent (grade 0).
   Example 2. Educational status: 0; 1-6; 7-8; 9 -12; more than 12.

3. Interval scale: This has the additional quality that the intervals between classes are equal.

   Example : Temperature (in Celsius)

   Equal differences between any pair of numbers in the scale indicate equal differences in the attribute being measured. The difference in temperature between 20 } C and 25 } C is the same as the difference between 30 } C and 35 } C. The ratio between numbers in the scale is not, however, necessarily the same as that between the amounts of the attribute. That is, a room at 30 } C is not 'twice as hot' as one at 15 } C. This is because the zero on the scale does not indicate absence of the attribute.

4. Ratio scale: This has the additional quality that zero indicates absence of the attribute. As a result, the ratio between numbers in the scale is the same as that between the amounts of the attribute being measured.

The existence of an adequate and up-to-date sampling frame often defines the study population.

Sampling methods

An important issue influencing the choice of the most appropriate sampling method is whether a sampling frame is available, that is, a listing of all the units that compose the study population.

a) Non-probability sampling methods

Examples:

1. Convenience sampling: is a method in which for convenience sake the study units that happen to be available at the time of data collection are selected.
2. Quota sampling: is a method that insures that a certain number of sample units from different categories with specific characteristics appear in the sample so that all these characteristics are represented. In this method the investigator interviews as many people in each category of study unit as he can find until he has filled his quota.
3. Purposeful sampling strategies for qualitative studies: Qualitative research methods are typically used when focusing on a limited number of informants, whom we select strategically so that their in-depth information will give optimal insight into an issue about which little is known. This is called purposeful sampling.

The above sampling methods do not claim to be repr esentative of the entire population.

Random sampling strategies to collect quantitative data: If the aim of a study is to measure variables distributed in a population (e.g., diseases) or to test hypotheses about which factors are contributing significantly to a certain problem, we have to be sure that we can generalise the findings obtained from a sample to the total study population. Then, purposeful sampling methods are inadequate, and probability or random sampling methods have to be used.

b) Probability sampling methods: They involve random selection procedures to ensure that each unit of the sample is chosen on the basis of chance. All units of the study population should have an equal or at least a known chance of being included in the sample.

1. Simple Random Sampling (SRS): This is the most basic scheme of random sampling. To select a simple random sample you need to:

   - Make a numbered list of all the units in the population from which you want to draw a sample. Each unit on the list should be numbered in sequence from 1 to N (Where N is the Size of the population).
   - Decide on the size of the sample
   - Select the required number of sampling units, using a “lottery” method or a table of random numbers.

2. Systematic Sampling: Individuals are chosen at regular intervals (for example, every 5^{th}, 10^{th}, etc.) from the sampling frame. Ideally we randomly select a number to tell us where to start selecting individuals from the list. For example, a systematic sample is to be selected from 1000 students of a school. The sample size is decided to be 100. The sampling fraction is: 100/1000 = 1/10. The number of the first student to be included in the sample is chosen randomly by picking one out of the first ten pieces of paper, numbered 1 to 10. If number 5 is picked, every tenth student will be included in the sample, starting with student number 5, until 100 students are selected. Students with the following numbers will be included in the sample: 5,15, 25, 35,45, . . . , 985, 995.

- ƒ Systematic Sampling is usually less time consuming and easier to perform than SRS.
- ƒ It provides a good approximation to SRS.
- ƒ Should not be used if there is any sort of cyclic pattern in the ordering of the subjects on the list.
- ƒ Unlike SRS, systematic sampling can be conducted without a sampling frame (useful in some situations where a sampling frame is not readily available).

4. Stratified sampling: If it is important that the sample includes representative groups of study units with specific characteristics (for example, residents from urban and rural areas), then the sampling frame must be divided into groups, or strata, according to these characteristics. Random or systematic samples of a predetermined size will then have to be obtained from each group (stratum). This is called stratified sampling.

Some of the reasons for str

6. Multi-Stage Sampling: This method is appropriate when the population is large and widely scattered. The number of stages of sampling is the number of times a sampling procedure is carried out.

   - The primary sampling unit (PSU) is the sampling unit (or unit of selection in the sampling procedure) in the first sampling stage;

- The secondary sampling unit (SSU) is the sampling unit in the second sampling stage, etc.

e.g. After selection of a sample of clusters (e.g. household), further sampling of individuals may be carried out within each household selected. This constitutes two-stage sampling, with the PSU being households and the SSU being individuals.

Advantages : less costly, we only need to draw up a list of individuals in the clusters actually selected, and we can do that when we arrive there.

Disadvantage : less precise than SRS.

When we take a sample, our results will not exactly equal the correct results for the whole population. That is, our results will be subject to errors. This error has two components: sampling and non-sampling errors.

a) Sampling error (i.e., random error)

Random error, the opposite of reliability (i.e., Precision or repeatability), consists of random deviations from the true value, which can occur in any direction.

Sampling error (random error) can be minimized by increasing the size of the sample.

Reliability (or precision): This refers to the repeatability of a measure, i.e., the degree of closeness between repeated measurement of the same value. Reliability addresses the

The sources of variation resulti ng in poor reliability include:

a) Variation in the characteristic of the subject being measured. Example: blood pressure
b) The measuring instruments, e.g. questionnaires
c) The persons collecting the information (observer variation)

Inter-observer variation: differences between observers in measuring the same observation

Intra-observer variation: differences in measuring the same observation by the same observer on different occasions.

b) Non Sampling error (i.e., bias)

Bias, the opposite of validity, consists of systematic deviations from the true value, always in the same direction.

It is possible to eliminate or reduce the non-sampling error (bias) by careful design of the sampling procedure.

Validity: This refers to the degree of closeness between a measurement and the true value of what is being measured. Validity addresses the question, how close is the measured value to the true value ?

To be accurate, a measuring device must be both valid and reliable. However, if one cannot have both, validity is more important in situations when we are interested in the absolute value of what is being measured. Reliability on the other hand is more important when it is not essential to know the absolute value, but rather we are interested in finding out if there is a trend, or to rank values.

Examples of types of bias in sampling include:

Bias resulting from incompleteness of the sampling frame: accessibility bias, seasonability bias, self-reporting bias, volunteer bias, non-response bias etc.

Non-response bias refers to failure to obtain information on some of the subjects included in the sample to be studied. It results in significant bias when the following two situations are both fulfilled.

1. When non-respondents constitute a significant proportion of the sample.
2. When non-respondents differ significantly from respondents.

The issue of non-response s hould be considered during the planning stage of the study:

a) Non-response should be kept to a minimum. E.g. below 15%

Methods that may help in maintaining non-response at a low level could be:

- Training data collectors to initiate contact with study subjects in a respectful way and convince them about the importance of the given study (this minimizes the refusal type of non-response)
- Offering incentives to encourage participation (this should be done by taking account of the potential problems that may arise in conducting future research)
- By making repeated attempts (at least 3 times) to contact study subjects who were absent at the time of the initial visit.

b) The number of non-responses should be documented according to type, so as to facilitate an assessment of the extent of bias introduced by non-response.

c) As much information as possible should be collected on non-respondents, so as to see in what ways they may differ from respondents.

- Selection bias cannot be corrected by increasing the size of the sample, why? How do you remove this type of bias?

6.7 Sample size determination

- In planning any investigation we must decide how many people need to be studied in order to answer the study objectives. If the study is too small we may fail to detect important effects, or may estimate effects too imprecisely. If the study is too large then we will waste resources.

- In general, it is much better to increase the accuracy of data collection (by improving the training of data collectors and data collection tools) than to increase the sample size after a certain point.

- The eventual sample size is usually a compromise between whTm 9r eusirabple and

Research methodology

I. Logistics of data collection

WHO will collect WHAT data?

When allocating tasks for data collection, it is recommended that you first list them. Then you may identify who could best implement each of the tasks. If it is clear beforehand that your research team will not be able to carry out the entire study by itself, you might plan to look for research assistants to assist in relatively simple but time-consuming tasks.

HOW LONG will it take to collect the data for each component of the study?

Step 1: Consider:

- The time required to reach the study area(s);
- The time required to locate the study units (persons, groups, records); If you have to search for specific informants (e.g., users or defaulters of a specific service) it might take more time to locate informants than to interview them.
- The number of visits required per study unit. For some studies it may be necessary to visit informants a number of times, for example if the information needed is sensitive and can only be collected after informants are comfortable with the investigator or if observations have to be made more than once (for example, follow-up of pregnant mothers or malnourished children). Time needed for follow-up of non-respondents should also be considered.

Step 2: Calculate the number of interviews that can be carried out per person per day

Step 3: Calculate the number of days needed to carry out the interviews. For example:

- you need to do 200 interviews,
- your research team of 5 people can do 5 x 4 = 20 interviews per day,
- you will need 200:20 = 10 days for the interviews.

Step 4: Calculate the time needed for the other parts of the study, (for example, 10 days)

Step 5: Determine how much time you can devote to the study.

If the team has fewer days for fieldwork than the required, they would need additional research assistants to help complete this part of the study.

Note:

It is always advisable to slightly overestimate the period needed for data collection to allow for unforeseen delays.

WHEN should the data be collected?
The type of data to be collected and the demands of the project will determine the actual time needed for the data to be collected. Consideration should be given to:

- availability of research team members and research assistants,
- the appropriate season(s) to conduct the field work (if the problem is season-related or if data collection would be difficult during certain periods),
- accessibility and availability of the sampled population, and
- public holidays and vacation periods.

II. Ensuring quality

It is extremely important that the data we collect are of good quality, that is, reliable and valid. Otherwise we will come up with false or misleading conclusions.

Measures to help ensure good quality of data:

- Prepare a field work manual for the research team as a whole, including:
  - — Guidelines on sampling procedures and what to do if respondents are not available or refuse to co-operate,
  - — A clear explanation of the purpose and procedures of the study which should be used to introduce each interview, and
  - — Instruction sheets on how to ask certain questions and how to record the answers.
- Select your research assistants, if required, with care. Choose assistants that are:
  - — from the same educational level;

— knowledgeable concerning the topic and local conditions;

— not the object of study themselves; and

— not biased concerning the topic (for example, health staff are usually not the best possible interviewers for a study on alternative health practices).

- Train research assistants carefully in all topics covered in the field work manual as well as in interview techniques and make sure that all members of the research team master interview techniques such as:

  — asking questions in a neutral manner;

  — not showing by words or expression what answers one expects;

  — not showing agreement, disagreement or surprise; and

  — recording the answers precisely as they are provided, without sifting or interpreting them.

- Pre-test research instruments and research procedures with the whole research team, including research assistants.
- Take care that research assistants are not placed under too much stress (requiring too many interviews a day; paying per interview instead of per day).
- Arrange for on-going supervision of research assistants. If, in case of a larger survey, special supervisors have to be appointed, guidelines should be developed for supervisory tasks.
- Devise methods to assure the quality of data collected by all members of the research team. For example, quality can be assured by:

  — requiring interviewers to check whether the questionnaire is filled in completely before finishing each interview;

— asking the supervisor to check at the end of each day during the data collection period whether the questionnaires are fille

Advantages:

- Documents can provide ready made information relatively easily
- The best means of studying past events.

Disadvantages:

- Problems of reliability and validity (because t

Example:

- to students of a school
- they can also be sent by post unlike interviews.

However, they demand a certain level of education on the part of the respondent.

On the other hand, interviews have many advantages:

- A good interviewer can stimulate and maintain the respondents interest
  the frank answering of questions.
- If anxiety is aroused (e.g., why am I being asked these questions?), the interviewer can allay it.

- An interviewer can repeat questions which are not understood, and give standardized explanations where necessary.

- An interviewer can ask “follow-up” or “probing” questions to clarify a response.

- An interviewer can make observations during the interview; i.e.,, noteg., o5/srod, an

In questionnaire design remember to:

a) Use familiar and appropriate language
b) Avoid abbreviations, double negatives, etc.
c) Avoid two elements to be collected through one question
d) Pre-code the responses to facilitate data processing
e) Avoid embarrassing and painful questions
f) Watch out for ambiguous wording
g) Avoid language that suggests a response
h) Start with simpler questions
i) Ask the same question to all respondents
j) Provide other, or don't know options where appropriate
k) Provide the unit of measurement for continuous variables (years, months, kgs, etc)
l) For open ended questions, provide sufficient space for the response
m) Arrange questions in logical sequence
n) Group questions by topic, and place a few sentences of transition between topics
o) Provide complete training for interviewers
p) Pretest the questionnaire on 20-50 respondents in actual field situation
q) Check all filled questionnaire at field level
r) Include "thank you" after the last question

Importance of combining differ ent data-collection techniques

A skillful use of a combination of different data-collection techniques can maximize the quality of the data collected and reduce the chance of bias. Investigators often use a combination of flexible and less flexible research techniques.

Flexible techniques, such as, loosely structured interviews using open-ended questions and focus group discussions are called qualitative research techniques . They produce qualitative information, which is often recorded in narrative form.

Structured questionnaires that enable the researcher to quantify pre- or post-categorized answers to questions are an example of quantitative research techniques . The answers to questions can be counted and expressed numerically.

Both qualitative and quantitative research techniques are often used within a single study.

Methods of collecting qualitative data

Qualitative approaches to data collection usually involve direct interaction with individuals on a one to one basis or in a group setting. Data collection methods are time consuming and consequently data is collected from smaller numbers of people than would usually be the case in quantitative approaches such as the questionnaire survey. The benefits of using these approaches include richness of data and deeper insight into the phenomena under study.

Unlike quantitative data, raw qualitative data cannot be analysed statistically. The data from qualitative studies often derives from face-to-face interviews, focus groups or observation and so tends to be time consuming to collect. Samples are usually smaller than with quantitative studies and are often locally based. Data analysis is also time consuming and consequently expensive.

The main methods of collecting qualitative data are: individual interviews, focus groups and observation

Qualitative interviews

Qualitative interviews are semi structured or unstructured. If the interview schedule is too tightly structured this may not enable the phenomena under investigation to be explored in terms of either breadth or depth. Semi structured interviews tend to work well when the interviewer has already identified a number of aspects he wants to be sure of addressing. The interviewer can decide in advance what areas to cover but is open and receptive to unexpected information from the interviewee. This can be particularly important if a limited

Focus group discussion

Sometimes it is preferable to collect information from groups of people rather than from a series of individuals. Focus groups can be usef

essentially quantitative studies where information could be more efficiently collected through structured interviews or questionnaires. Conversely, self completed questionnaires are

unanswered (or new questions which come up) can be addressed before data collection is over.

Preparation of a plan for data processing and analysis will provide you with better insight into the feasibility of the analysis to be performed as well as the resources that are required. It also provides an important review of the appropriateness of the data collection tools for collecting the data you need. That is why you have to plan for data analysis before the pre-test. When you process and analyze the data you collect during the pre-test you will spot gaps and overlaps which require changes in the data collection tools before it is too late!

What should the plan include?

When making a plan for data processing and analysis the following issues should be considered:

- Sorting data,
- Performing quality-control checks,
- Data processing, and
- Data analysis.

Sorting data

An appropriate system for sorting the data is im

It is useful to number the questionnaires belonging to each of these categories separately right after they are sorted.

> For example , the questionnaires administered to users of family planning services could be numbered U1, U2, U3, etc., and those for the non-users N1, N2, N3, etc.

Performing quality control checks

Usually the data have already been checked in the field to ensure that all the information has been properly collected and recorded. Before and during data processing, however, the information should be checked again for completeness and internal consistency.

If a questionnaire has not been filled in completely you will have MISSING DATA for some of your variables. If there are many missing data in a particular questionnaire, you may decide to exclude the whole questionnaire from further analysis.

- If an inconsistency is clearly due to a mistake made by the researcher/research assistant (for example if a person in an earlier question is recorded as being a non-smoker, whereas all other questions reveal that he is smoking), it may still be possible to check with the person who conducted the interview and to correct the answer.
- If the inconsistency is less clearly a mistake in recording, it may be possible (in a small scale study) to return to the respondent and ask for clarification.
- If it is not possible to correct information that is clearly inconsistent, you may consider excluding this particular part of the data from further processing and analysis as it will affect the validity of the study. If a certain question produces ambiguous or vague answers throughout, the whole question should be excluded from further analysis. (Normally, however, you would discover such a problem during the pre-test and change the wording of the question.)

For computer data analysis, quality control checks of data must also include a verification of how the data has been transformed into codes and subsequently entered into the computer. The same applies if data are entered into master sheets.

2. Coding

If the data will be entered in a computer for subsequent processing and analysis, it is essential to develop a CODING SYSTEM.

For computer analysis, each category of a variable can be coded with a letter, group of letters or word, or be given a number. For example, the answer ‘yes’ may be coded as ‘Y’ or 1; ‘no’ as ‘N’ or 2 and ‘no response’ or ‘unknown’ as ‘U’ or 9.

The codes should be entered on the questionnaires (or checklists) themselves. When finalising your questionnaire, for each question you should insert a box for the code in the right margin of the page. These boxes should not be used by the interviewer. They are only filled in afterwards during data processing. Take care that you have as many boxes as the number of digits in each code.

Coding conventions

Common responses should have the same code in each question, as this minimises mistakes by coders.

For example:

Yes (or positive response)	code - Y or 1
No (or negative response)	code - N or 2
Don’t know	code - D or 8
No response/unknown	code - U or 9

Codes for open-ended questions (in questionnaires) can be done only after examining a sample of (say 20) questionnaires. You may group similar types of responses into single categories, so as to limit their number to at most 6 or 7. If there are too many categories it is difficult to analyze the data.

2. Summarizing the data in data master sh eets, manual compilation, or compilation by computer

(1) Data master sheets

If data are processed by hand, it is often most efficient to summaries the raw research data in a so-called DATA MASTER SHEET, to facilitate data analysis. On a data master sheet all the answers of individual respondents are entered by hand.

To illustrate the use of master sheets, we will give an example of a rapid appraisal carried out by students of a nursing school about the smoking habits of the inhabitants of their town. The questionnaire had only 17 questions, of which 9 were asked of everyone, 4 exclusively to smokers and 4 exclusively to non-smokers. It was therefore decided to process the data by hand, divided in two groups: smokers and non-smokers, which were again subdivided in males and females. For each of the four groups, master sheets were prepared, on which all the answers of individual respondents could be recorded.

Master sheets can be made in different ways. For short simple

ii. Data entry

To enter data into the computer you have to develop a data entry format, depending on the program you are using. After deciding on a data entry format, the information on the data collection instrument will have to be coded (e.g., Male: M or 1, Female: F or 2). During data entry, the information relating to each subject in the study is keyed into the computer in the form of the relevant code.

iii. Verification

During data entry, mistakes will definitely creep in. The computer can print out the data exactly as it has been entered, so the printout can be checked visually for obvious errors, (e.g., exceptionally long or short lines, blanks that should not be there, alphabetic codes where numbers are expected, obviously wrong codes).

Example:

- Codes 3-8 in the column for sex where only 1(F) and 2 (M) are possible
- Codes above 250 when you had only 250 subjects

If possible, computer verification should be built in. This involves giving the appropriate commands to identify errors.

Example:

The computer can be instructed to identify and print out all subjects where the ‘sex’ column has a code different from 1 (F) or 2 (M).

iv. Programming

A certain amount of basic knowledge of computer programming is needed to give the appropriate commands.

v. Computer outputs

The computer can do most of the analysis and the results can be printed. It is important to decide whether each of the tables, graphs, and statistical tests that can be produced makes sense and should be used in your report. That is why we PLAN the data analysis BEFOREHAND!

Data analysis – quantitative data

Analysis of quantitative data involves the production and interpretation of frequencies, tables, graphs, etc., that

appeared to be a critical value. The children in families with an income above average had had significantly less diarrhoea than the children in families with an income below average.

(3) Construct a table indicating how data are grouped and count the number of observations in each group.

2. Cross-tabulations

Further analysis of the data usually requires the combination of information on two or more variables in order to describe the problem or to arrive at possible explanations for it.

For this purpose it is necessary to design CROSS-TABULATIONS.

Depending on the objectives and the type of study, two major kinds of cross-tabulations may be required:

- Descriptive cross-tabulations that aim at describing the problem under study.
- Analytic cross-tabulations in which groups are compared in order to determine differences, or which focus on exploring relationships between variables.

When the plan for data analysis is being developed, the data, of course, are not yet available. However, in order to visualize how the data can be organized and summarized it is useful at

- All tables related to a certain objective should be numbered and kept together so the work can be easily organised and the writing of the final report will be simplified.

To further analyse and interpret t

Step 2

form part of a jigsaw puzzle that you have to put together in order to obtain insight in your problem/topic under study.

IN CONCLUSION, a plan for the processing and analysis of data may include:

- a decision on whether all or some parts of the data should be processed by hand or computer;
- dummy tables for the description of the problem, the comparison of groups (if applicable) and/or the establishment of relationships between variables, guided by the objectives of the study;
- a decision on the sequence in which tables or data from different study populations should be analysed;
- a decision on how qualitative data should be analysed;
- an estimate of the total time needed for analysis and how long particular parts of the analysis will take;
- a decision concerning whether

Non-Maleficence- we ought not to in flict evil or harm

This principle states that we may not inflict harm on or expose people to unnecessary risk as a result of our research project. This is particularly important if our subjects may not be competent in some way, such as, the ability to give informed consent.

Beneficence – we ought to further othe rs' legitimate interests

This is the principle that obliges us to take positive steps to help others pursue their interests. These interests clearly have to be legitimate.

Justice- we ought to ensure fair entitlement to resources

This principle is concerned with people receiving their due. This means people should be treated equally in every way since not all people are equally competent or equally healthy.

Ethical rules

The ethical rules of research, like principles, are not absolute in that one may override another although clearly this must be justified. These rules are essential for the development of trust between researchers and study participants. Like the ethical principles on which the rules are based, there are four:

Veracity

All subjects in any research project should always be told the truth. There is no justification for lying, but this is not the same non-disclosure of information should it, in particular, invalidate the research.

Privacy

When subjects enroll in a research study, they grant access to themselves, but this is not unlimited access. Access is a broad term and generally includes viewing, touch or having information about them.

Confidentiality

Although someone may grant limited access to him or herself, they may not relinquish control over any information obtained. Certainly, no information obtained with the patient's or

Pretesting is: 1. Simpler

2. Less time consuming

3. Less costly than conducting an entire pilot study.

Grade	Number of students	Number of sections
9	600	8
10	500	7
11	400	6
12	300	5
Total	1800	26

a) What type of sampling technique do you use? Why?

b) How do you select the subjects who will be included in your sample?

6. A multi-national clinical trial is proposed to investigate the value of a gradually increasing dose schedule of a beta blocker in the treatment of severe heart failure. The trial will be randomised, double-blind and placebo controlled. Each patient is to be followed for 2 years, and the main treatment comparison is for all cause mortality. Previous experience suggests a 2 year mortality rate of around 30%. The investigators propose that a one-third reduction in mortality due to beta-blockade would be important to detect. They suggest that type I and type II errors be set at .05 and .1, respectively.

   a) Calculate the required number of patients to be recruited.

   b) Suppose one anticipates that 10% of patients randomised to beta-blockade will fail to comply with the intended treatment policy. What change in required sample size would you suggest?

7. Prepare your data-collection tools, taking care that you cover all important variables of your proposed study.

Example of a GANTT Chart

Example of a budget proposal

	Budget Category	Unit Cost	Multiplying factor	Total Cost
				(Birr)
1.	Personnel	Daily Wage (including per diem)	Number of staff days (Number of staff x Number of working days)	
	Principal investigator	100	1x15	1,500.00
	Supervisors	100	2 x 15	3,000.00
	Data collectors	60	10 x 15	9,000.00
	Data entry clerk	40	1x 20	800.00
	Secretarial work	40	1x20	800.00
	Sub total		Personnel TOTAL	

CHAPTER EIGHT

- Why the proposed study is important, i.

- Operational definitions
- Some elements of the variables to be studied:

What characteristics will be measured? How will the variables be defined? What scales of measurement will be used etc.

ƒ Inclusion/ exclusion criteria

ƒ Sampling procedure to be used and sample size and power calculation.

ƒ Data collection and management

- Data collection and coding forms should be appended to protocol
- Training and quality control, bias control, data entry and storage, data clean-up and correction of deficiencies

ƒ Data analysis

- Management of dropouts
- Frequencies, rates, other parameters
- Statistical programs and tests to be used
- Data presentation (dummy tables to be appended)

ƒ Ethical considerations: rights

V) Budget (itemize all direct costs in Ethiopian Birr)

- Personnel, material/supplies, travel, analysis, contingency, etc.

VI) References: List only those cited in text and number by order they appear in text using Arabic numerals.

VII) Appendices: -

- Data collection and coding forms
- Dummy tables for data presentation
- Letters of support (cooperation)

Exercise

Develop a research proposal of your own topic. Take account of all the chapters covered so far and write your final proposal in line with the guideline given above.

8.2 Summary of the major activities of the fieldwork phase

Activities to be performed during the field period

1. Briefing of managers and health service personnel: The purpose of the briefing is to obtain support for the project. Such support is necessary to obtain resources as well as to get permission to collect the required data. Briefing should be conducted with all important persons and/or organizations at different levels.

2. Identifying and obtai ning project resources : We have to identify and obtain the necessary resources (manpower, materials, etc.) needed to collect the required data. We have to make sure that all the items needed for the study are ready.

3. Reviewing availability of subjects : It is important to make personal visits to every site where the actual data will be collected to understand the physical and manpower limitations, constraints and special circumstances that could influence data collection. This would assist the investigators to take an appropriate measure and make the necessary preparations.

4. Organizing logistics for data collection: Having made an inventory of available resources, the logistics for data collection have to be organized. This will involve planning in detail how , where, and when data collection will be carried out.

5. Preparing fieldwork manuals: Manuals or instruction sheets should be prepared for

pretest, they could participate in the adjustment of instruction sheets and data-collection tools.

During the training, the data colle

8.3 Writing a research report

Writing a good report may take much time and effort. The most difficult task is usually the preparation of the first draft. The report should be easily intelligible. This requires clarity of language, a logical presentation of facts and inferences, the use of easily understood tables and charts, and an orderly arrangement of the report as a whole. It should be no longer than is necessary.

Conventionally, a report usually contains the following major components.

Title and cover page

The cover page should contain the title, the names of the authors with their titles and positions, the institution that is publishing the report, (e.g., Gondar College of Medicine and Health Sciences) and the month and year of publication. The title could consist of a challenging statement or question, followed by an informative subtitle covering the content of the study and indicating the area where the study was implemented.

Abstract (Summary)

The summary should be brief and informative. A reader who has been attracted by the title will usually look at the summary to decide whether the report is worth reading. The summary should be written only after the first or even the second draft of the report has been completed. It should contain:

— a very brief description of the problem (WHY this study was needed)

— the main objectives (WHAT has been studied)
— the place of study (WHERE)
— the type of study and methods used (HOW)
— major findings and conclusions, followed by
— the major (or all) recommendations.

The summary will be the first (and for busy health decision makers most likely the only) part of your study that will be read. Therefore, its writing demands thorough reflection and is time

Information that would be useful to special categories of readers but is not of interest to the average reader can be included in annexes as well.

Examples of information that can be presented in annexes are:

- tables referred to in the text but not included in order to keep the report short;

1. VDU (visual display unit): it works like a television because it allows you to see the work that you do with the computer.

2. Motherboard:

2. The central processing unit (CPU)
   All computations, regulation of data flow, decision-making operation, etc. are carried out by this part of the computer.

   The CPU is an important component of the computer system that contains the control unit, the arithmetic-logic unit (ALU), and main memory (Primary storage).

   All these parts work together to electronically control the functions of the computer system.

3. Auxiliary (secondary) storage devices:
   To store large amount of in

A typical data set will consist of a collection of cases (or records), each of which contains the values of a set of variables (or fields).

Although it would be possible to write computer programs to organize and analyse a file of data, in general Database management systems (DBMSS) are now used. These are ready-written, general purpose programs to help in the organization and analysis of data. Two examples are: Epi-Info and SPSS.

I. Introduction to Epi Info 6

x Epi Info is a multi-purpose computer program designed for epidemiological researchers.

Helpful Tips

a) File names should be recognizable ones and should not exceed eight characters. A character could be a letter or a number.

b) The extension for an Epi Info questionnaire file name should always be "qes".

Example

birthwt.qes

trachoma.qes

Exercise :

Develop a questionnaire of your own data and prepare a template using the EPED program.

Entering Data using the ENTER program

- The ENTER program will create a data file from a questionnaire. That is, The ENTER program will create an Epi Info database < .rec > file using the questionnaire.

- Once the < .rec > file is created, the file may be loaded into ENTER for adding more records or editing those already entered.

- If the questionnaire is revised, ENTER can be instructed to revise the < .rec > file accordingly.

- In the ENTER module, there are five options to choose from. We will see the first three ones which are in common use.

    1) Enter or edit data

    2) Create new data file from .Qes file

    3) Revise structure of data file using revised .Qes

- Now, move the cursor to the ENTER module and press the Enter key.

- Write the file name on the space given below "Data file < .Rec> : "

- Write the numbers 1 or 2 or 3

- Press F1
- Write the maximum number (e.g., the maximum number for women aged 15 to 49 years is 49)

- Press F2
- Finally, the valid values will be indicated at the bottom of your questionnaire. In this case, the valid values will be 15 to 49.

- Save data

b) F7 - Jump

Appropriate jumps can be built in so that questions which are not applicable as a result of a previous response are jumped.

- put the cursor on the blank spaces of any of the fields (variables).
- Write any number different from the valid ones

- Press F7
- Pu

- Move the cursor to "ANALYSIS of data " and press the Enter key.
- Use the "READ" command to choose a dataset.

  Example : Read birthwet.rec
- Once your .Rec file is retrieved, you can perform the tasks explained above.

  Example

  freq age
  freq sex
  tables age sex
  List sex
  Update age sex etc.
- If you press the function key "F2", you will see the various "commands" that perform different tasks.
- If you press the function key "F3", you will see the list of "variables" that you have created using the EPED program.
- If you press the function key "F5", your output will go to the printer. If you press it again, the output will go to the screen.

DEFINE and RECODE - These commands are very important in facilitating the analysis of data.

- DEFINE allows creation of new variables for use in analysis.
- Variable names in the DEFINE statement must be 10 or fewer characters.
- Variable names do not begin with a number.

Example :

DEFINE age1 ## (a newly created variable)

This could be used in a RECODE statement to provide labels for a numerically coded age variable.

- The RECODE command is used to form several categories (groups).

Example :

RECODE variable1 to variable2 15-19=1 20-24=2 25-29=3 30-34=4 35-39=5 40-44=6 45-49=7

- Variable1 refers to the old variable while variable2 refers to the newly created one.
- The variable "age" given earlier is an old one. Therefore, we can put it in place of variable1
- The variable "age1" shown above is a new one. Therefore, we can put it in place of variable2.

After successfully completing the recoding of variables, you can use other commands (e.g., freq, tables, list, etc.) to have a summary statistics based on the new variables.

Helpful tips

a) The update command will assist you to see all the data values you have entered. You can make any changes using this command.

b) Pressing F10 will take you one step back.

STATCALC calculator

- STATCALC does statistical analysis of data entered from the keyboard into tables on the screen.
- Facilities are provided to perform 2 by n tables , to investigate

Example:

A cross-sectional survey on knowledge and use of condom was carried out among commercial sex workers (CSWs) of three small towns in Northwest Ethiopia. The table below shows the experience of these women on the use of condom classified by some selected

Variable	study population (n = 6008)	Number of persons who underwent THHP* (n = 368)
Age group (in years)		
< 1	194	41
1 - 4	1255	56
5 - 14	1829	58
15 - 44	2381	184
45 - 64	312	25
65+	37	4
Sex		
Male	2992	195
Female	3016	173

* = traditional harmful health practice

Investigate the impact of age and sex on the experience of the study subjects towards THHPs.

Epi6 performs a lot of tasks using simple statistical techniques. However, it should be noted that Epi6 does not have facilities to perform multivariate analysis, such as, logistic regression. The multiple linear regression technique contained by thariate a5t5 0 T0.0002 - -02(]TJ 15.14

2) The question being asked must have a "yes/no" or other two-choice answer, leading to a proportion of the population <the "yes's" > as the final result.

Example : Suppose that you wish to investigate whether or not the true prevalence of HIV antibody in a population is 10%. You plan to take a random sample of the population to estimate the prevalence. You would like 95% confidence interval that the true proportion in the entire population will fall within the confidence intervals calculated from your sample.

In STATCALC, therefore, you enter the population size, say 5000, the estimate of the true prevalence (10%), and either 6% or 14% as the "worst acceptable" value, the end point of your sample confidence interval. The program then shows the sample size for several different confidence levels, including the 95% we desired.

The "worst acceptable" value is one of the confidence limits around the estimated sample proportion. The sample size given is for a "two-tailed test", a larger sample size than for a "one-tailed test". The equation works with the following values, if the confidence interval is 95%.

Upper worst acceptable = proportion (P) + (1.96 standard error) = P + marginal error
Lower worst acceptable = proportion (P) - (1.96 standard error) = P - marginal error

In the sample size calculations, an initial screen explains the data items and allows input of a set of values. Pressing <F4> then shows the results (calculated sample sizes) on the second screen.

Sample size calculations for different study designs (more complex designs) are also provided by the STATCALC program.

II. Introduction to SPSS

This section was taken from a resource pack entitled "An Introduction to Practical Statistics Using SPSS" , (Tent Focus, 2002).

The Data Editor window is similar to a spreadsheet. The rows represent individual cases (observations) and the columns represent variables[2]. A single cell is an intersection of a case and a variable e.g. the height of person x.
The Output window is where SPSS displays the statistics and reports from the analysis of your data.

Entering Data

When you start SPSS you are automatically placed in the Data Editor window. The active cell in the window has a heavy black border around it, indicating that any data you type will be entered into that cell. You can move around the Data Editor window by using the arrow keys (8 , 9 : , ;), or by clicking on cells with the mouse.

Table1 presents some patient data that we can enter in to SPSS. For each patient we have collected the following data from their medical notes: gender, age and blood group. The data has already been coded[3] by a researcher.

Table1: Patient data

Gender	Age	Blood group
1	21 2	
2	39 1	
2	43 1	
1	55 1	
1	26 4	
1	19 4	
2	65 2	
2	41 2	
1	61 3	
1	50 1	

Where for Gender: 1 = male, 2 = female
and for Blood Group: 1 = O, 2 = A, 3 = AB, 4 = B.

Decimals indicates the number of decimal places to be shown

. Click on in the first cell of the fourth column

. Type 0 in this cell, or click on the down arrow twice to reduce the number shown to 0

Before continuing examine the range of data types that are available. Clicking on any cell in the

Type column and then clicking on the …

Option 2: User-defined missing values

Here codes defined by the user are declared as missing values, using the missing option in the define missing values dialogue (this is accessed via the define variable dialogue).

Once all of the variables are correctly defined, clicking on the Data View tab at the bottom of the window will return you to the data sheet.

Research methodology

To save the data

. Click on the File menu

. Click on Save As…

. Once you have chosen the co

Research methodology

- Do the same thing until all the required independent variables are considered.

x Click on OK

x Move the cursor to Statistics

- Click on cancel
- Click on File
- Move the cursor to open and click on Data
- Look in: SPSS (press the down arrow and click on Local disk (c:)
- move the cursor to Epi6 and click on open
- Files of type: (press the down arrow and click on dbase (*.dbf)
- Now, the required file appears below the EPI directory and click on it.
- click on open
- At this stage, you should see some descriptions regarding the type of variables and the number of cases (records)
- Move the cursor to Untitled SPSS Data and click on it.
- Delete any unnecessary records at the first row (A1) (if any)
- Finally, save the file as an SPSS data file with extension .sav

REFERENCES

1. Corlien M. Varkevisser, Indra Pathmanathan, and Ann Brownlee. Designing and conducting health systems research projects: Volume 1 Proposal development and fieldwork. KIT/IDRC. 2003
2. Degu G, Tessema F. Biostatistics for Health Science Students: lecture note series. The Carter Center 9EPHTI), Addis Ababa; January 2005.
3. Abramson JH. Survey methods in community medicine. 2nd ed.
   Eidenburgh: Churchill Livingstone, 1979.
4. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall, 1991.
5. Colton T. Statistics in Medicine. Boston: Little, Brown and Company 9INC.),
   November 19974.
6. Mathers, Nigel; Howe, Amanda; and Hunn Amanda. Trent focus for research and development in primary health care. Ethicamodules [(5h-4.4(t in siderationAba4edi[illegible]