Research Methods for Promotion of Lung Health

A guide to protocol development for low-income countries

2001
International Union Against Tuberculosis and Lung Disease

Donald A Enarson MD
Susan M Kennedy PhD
David L Miller MD
Per Bakke MD PhD

From materials developed for
International Courses on Research Methods for the Promotion of Lung Health
Sponsored by the IUATLD
Publication supported by a grant from USAID

International Union Against Tuberculosis and Lung Disease
68 boulevard Saint-Michel, Paris 75006 FRANCE

Preface

The International Union Against Tuberculosis and Lung Disease (IUATLD) is the oldest international non-governmental organisation dealing with health in the world. Its specific role is to develop strategies and mechanisms to promote lung health and to prevent lung disease. Its focus is on the most frequent and serious problems affecting the lung and concentrates its activities toward low-income countries where most of those at risk live.

To carry out its work, the IUATLD has adopted research as one of its priority activities. It has done this because it sees research as one of the three pillars upon which to carry out effective action, the other two being technical assistance to public health programs and education activities. The Commission on Health Research for Development[1], an initiative whose goal has been to improve the health of people in low-income countries, concluded that research "has enormous – and, in great part, neglected – power to accomplish that goal" (improving health). It went on to point out that reaching the goal depends on commitment to the goal, on programs to reach people, on political will to provide resources to carry them out and on knowledge to guide actions. Knowledge is generated by research and is "essential for effective action for health".

The experience of the IUATLD, through the International Respiratory Diseases Research Unit, has identified certain basic requirements for the effective conduct of a research program in low-income countries[2]:

1. structures mandated to carry it out;
2. competent personnel to do it;
3. interaction of researchers to stimulate the relevant intellectual climate.

(1) Evans JR, Castillo GT, Abed FH, et al. Health research: essential link to equity in development. New York: Oxford University Press, 1987; 1-136.

(2) Becklake MR. Report of the International Respiratory Diseases Research Unit. Paris: IUATLD, 1993.

To meet these requirements, the IUATLD, in collaboration with organisations sharing its goals, has developed an international network of courses on research methods for the promotion of lung health. The stated aims of this initiative were:

1. to promote research within institutions in low-income countries;
2. to build the training capacity for basic research in these institutions;
3. to expand expertise in more advanced research techniques;
4. to develop materials to be used by the institutions for ongoing training;
5. to encourage international collaborative research by members in low-income countries;
6. to foster research "partnerships" among institutions.

The material in this book addresses aim number 4. It was developed initially for use in the course on Research Methods for Promotion of Lung Health held in Istanbul in 1997. The textbook, "Lecture notes on epidemiology and public health medicine"[3] was used in the course, and the materials built on the experiences gained in similar IUATLD courses in other locations.

This book contains two parts. The initial section comprises course notes and the second, practical exercises and instructions on using the computer package, Epi Info in the research that this volume seeks to promote. This material is available both in printed form and on the IUATLD website (www.iuatld.org). The practical exercises can be expanded and/or adapted for local use.

The courses have been designed to assist health workers and investigators in developing research protocols relevant to the actual situation in low-income countries. The materials follow a specific course format consisting of four distinct categories of material: lectures on basic epidemiology, practical exercises prepared to illustrate these lectures, sessions on data management using Epi Info, sessions on protocol development.

The contents of these various components have been enriched by contributions from other international faculty in Istanbul and other courses, and by our interaction with participants. Their use by our colleagues in future courses will undoubtedly serve to further amplify and improve them. In addition, it is hoped that they will be used for teaching in local institutions in low-income countries throughout the world. They will need to be modified to address the specific needs of particular groups and to be appropriate to the local situation.

(3) Farmer R, Miller D, Lawrenson R. Lecture notes on Epidemiology and Public Health Medicine, 4th ed. Oxford: Blackwell Science. 1995.

Table of contents

1. PROMOTING LUNG HEALTH 9

1.1 What does it mean to promote lung health? 9

1.1.1 Understanding health
1.1.2 What is lung health?
1.1.3 How can we promote lung health?

1.2 What is research and why is it important? 11

1.3 How are research priorities established? 11

1.3.1 How frequent is the condition relative to other conditions?
1.3.2 What is the degree of disability or dysfunction due to the condition?
1.3.3 Are there cost-effective means to cure, control or prevent lung diseases?

2. EPIDEMIOLOGY, A BASIC SCIENCE FOR LUNG HEALTH 15

2.1 What is epidemiology and how is it used? 15

2.1.1 History
2.1.2 Definitions
2.1.3 The nature of epidemiological research

2.2 Applications of epidemiology 18

2.2.1 Measuring the health of a population
2.2.2 Describing the natural history of disease
2.2.3 Discovering the determinants of disease
2.2.4 Controlling and preventing disease
2.2.5 Planning and evaluating health services

2.3 Understanding association and causation 20

2.3.1 What is Association?
2.3.2 What is Causation?

3. GETTING STARTED IN RESEARCH 23

3.1 The research protocol 23

3.1.1 Why focus on the research protocol?
3.1.2 Typical format and elements of the research protocol

3.2 Deciding what to study: The research question 32

3.2.1 Basic concepts in framing a research question
3.2.2 Points to consider when choosing a research question

4. STRUCTURING RESEARCH: STUDY DESIGN 35

4.1 Study design: its relation to the development of medical knowledge 35

4.2 Overview of study architecture 37

4.3 Study design details 37

4.3.1 Descriptive studies
4.2.2 Analytic studies
4.2.3 Experimental (intervention) studies

5. THE SUBJECT OF RESEARCH: SELECTING A POPULATION 51

5.1 Defining a population to study 51

5.2 Selecting a sample 52

5.2.1 Definitions and procedures
5.2.2 How do I select the sample?
5.2.3 How do I ensure that wrong conclusions are not drawn?

5.3 Deciding how big the study population should be 55

5.3.1 Why is the size of the study important?
5.3.2 What information is needed to determine the size of the study?
5.3.3 How do I do a sample size calculation using these factors?

6. MEASUREMENT IN EPIDEMIOLOGY 63

6.1 General principles 63

6.1.1 Principles of collecting information (data)
6.1.2 Types of data
6.1.3 Instruments to measure states or characteristics (collect data)
6.1.4 Problems with measurement

6.2 Counting diseases and measuring health 67

6.2.1 Sources of information
6.2.2 Studies of mortality
6.2.3 Studies of morbidity

6.3 Use and design of a questionnaire 73

6.3.1 Principles of questionnaire design
6.3.2 Content of the questionnaire

6.4 Measuring exposure 77

6.4.1 Recording or measuring environmental exposure
6.4.2 Instruments for collecting exposure information
6.4.3 Choosing an exposure assessment method

6.5 Reporting morbidity or mortality 81

6.5.1 Calculation of rates
6.5.2 Indices of risk

7. CONDUCTING RESEARCH PRACTICAL STEPS 85

7.1 Project Execution: your protocol has been accepted for funding – what now? 85

7.1.1 The research plan
7.1.2 Controlling the quality of procedures and techniques
7.1.3 Ensuring efficiency

7.2 Managing the information collected 89

7.2.1 Checking original data collection forms and records
7.2.2 Coding data
7.2.3 Entering data into computer files
7.2.4 Checking computer files
7.2.5 Handling missing data

8. INTERPRETING RESULTS 97

8.1 Ensuring that the results are valid 97

8.1.1 Definitions
8.1.2 Error and bias in sampling study populations (selection biases)
8.1.3 Error and bias in measurement (information biases)
8.1.4 Assessment of the error inherent in a test or instrument
8.1.5 Confounding (a special kind of selection bias)

8.2 Making sense of the results - Data Analysis 105

8.2.1 Getting acquainted with the data
8.2.2 Approaches to statistical analysis
8.2.4 Statistical tests for comparing discrete variables between groups
8.2.5 Statistical tests relevant for continuous measures: compared across 2 or more groups
8.2.6 Statistical tests relevant for continuous measures: association between 2 measures
8.2.7 Statistical tests that consider comparisons, while "controlling" for other variables
8.2.8 Other issues in statistical analyses

8.3 Reporting the results 124

8.3.1 Preparing a report of the research
8.3.2 Writing a scientific paper

9. OTHER ISSUES IN RESEARCH 129

9.1 Who owns the research? 129

9.1.1 Structuring responsibility
9.1.2 Addressing perceived needs
9.1.3 Ensuring follow up of the results

9.2 Getting support for research 130

9.2.1 Local sources
9.2.2 International Sources

9.3 Ethical issues in research 133

9.3.1 Basic principles
9.3.2 Principles relevant to sponsored research in low-income countries
9.3.3 Appropriate structures for ethics review

CHAPTER 1

PROMOTING LUNG HEALTH

Research is essential to provide the knowledge necessary for action to improve the health of the community. Whatever the nature of health research, the ultimate aim is to improve the health of the individual and of the community.

1.1 What does it mean to promote lung health?

1.1.1 Understanding health

Before moving on to a discussion of research for promotion of lung health, we must develop a clear conception of health itself and what it means to promote it. In order to have clarity and consistency in the terms used, definitions for the terms will be taken from a single source, unless otherwise stated(4).

What is health and how is it measured? The World Health Organization defines health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity". This definition has been criticised as lacking precision and consequently being very difficult, if not impossible, to measure.

The subject of research in the context of this text includes diseases that cause morbidity and death, structure and utilisation of health services, and policies which may affect health.

- Disease is a physiological or psychological dysfunction.
- Health services are services by or under the direction of health care professionals for promoting, maintaining or restoring health.
- Policies refer to the complex combination of psychosocial and political issues affecting health, whether beneficial or adverse.

(4) Last JM ed. A Dictionary of Epidemiology 3rd ed. New York: Oxford University Press, 1995, pp 180.

1.1.2 What is Lung Health?

The lung is a large and vulnerable interface between each human being and the external environment. Its function is to extract oxygen from the environment, which is essential for maintaining life, and to expel gases whose retention would threaten life. The interface between the lung and the environment, therefore, must offer a minimal barrier to the transport of these gases. Anything that impairs the efficiency of the gaseous exchange is detrimental to lung function. This can occur in a variety of disease processes.

The major lung diseases result from exposure to harmful agents in the ambient environment. The two major groups of diseases are those due to exposure to micro-organisms (the most frequent being pneumonia and tuberculosis) and to particles, fumes and gases (the most frequent being chronic airflow obstruction, asthma, lung cancer and pulmonary fibrosis).

The prevention and management of lung diseases and the promotion of lung health depends on a clear understanding of the interaction between an individual and the many potentially harmful agents in the surrounding environment.

1.1.3 How Can We Promote Lung Health?

Health promotion is the process of enabling people to increase control over and improve their health through action directed at the determinants of health.

The vehicle for health promotion is the health service, both clinical and preventive. Consideration of the structure and efficiency of health services is, therefore, fundamental to our purposes.

Recent developments in the structure and operation of health services in many countries have led to an emphasis on technical efficiency, accountability and cost-savings. To obtain the best care, duly accounted for, at the lowest cost is the aim of the modern health service.

But, what is the best care? How can it be monitored to ensure that it remains good as it is implemented? Will cost-savings result in deterioration of the quality of the service? These are issues of key importance to all health care and public health workers today.

Health services for promoting, maintaining or restoring lung health are usually a component of general health services and rarely (except at the highest level of specialised services) are services dealing only with lung diseases. Investigations to develop or evaluate such services (health services research)consist of several areas of evaluation:

- Structure, concerned with resources, facilities and personnel,
- Process, concerned with where, by whom and how services are provided,
- Output, concerned with the amount and nature of the services provided,
- Outcome, concerned with the results.

1.2 What is research and why is it important?

The purpose of research is to create the knowledge essential for action to improve health. Without this knowledge, action is impossible because it has no logical or empirical basis. Indeed, ongoing action for health, if it does not contain an imbedded program of research, frequently becomes irrelevant, misleading or unnecessarily costly. It is for this reason that the original conception of the National Tuberculosis Program (a forerunner of many other public health programs) included the necessity of an imbedded program of research as an essential component.

Research is an activity of perpetual questioning. While public health practice is based on consensus, standardization and systematic practice, research requires a skeptical mind, prepared to continuously evaluate and question. This questioning and evaluating, when put into a systematic framework, creates the new knowledge that is required to create and continually modify actions for health. This is what research is and why it is important.

1.3 How are research priorities established?

In establishing priorities for research in the health sciences and within the health services, the following questions need to be addressed:

1.3.1 How frequent is the condition relative to other conditions?

Lung diseases are second only to the composite impact of all other infectious diseases as a burden to health globally(5). They account for over one-eighth of all disease and more than ten million people die each year from the most frequent forms of lung disease. These are:

- Acute respiratory infections, the single most frequent cause of death in children under the age of five years;
- Tuberculosis, the most frequent cause of death from a single agent in young adults between the ages of fifteen and forty-nine;
- Diseases caused by tobacco smoke, which account for as many deaths as tuberculosis and are implicated in more than one quarter of all deaths in many industrialised countries;
- Deaths from the effects of air pollution which reach millions in the large cities of Asia alone;
- Disabling airway disease (chronic airflow obstruction and asthma) from exposure to mineral and organic dusts, fumes, and sensitising chemicals, especially in newly emerging industries.

(5) World Bank. World Development Report 1993: Investing in health. World Bank, Oxford University Press, Oxford, 1993.

Death from acute respiratory infections and tuberculosis occur almost exclusively in low-income countries. Deaths from air pollution and workplace exposures are much more frequent in low-income countries. At present most deaths from smoking tobacco occur in industrialised countries, but the burden is rapidly increasing in low-income countries.

1.3.2 What is the degree of disability or dysfunction due to the condition?

Some lung diseases are not major causes of death but, due to their frequency or chronicity, cause much disability and impose heavy demands on health services.

- Acute respiratory infections are frequently among the top three reasons for health services utilisation at the primary care level;
- There are more than 100 million cases of asthma in the world today;
- Improper case management of asthma leads to huge costs from inappropriate use of health services;
- Improper case management of tuberculosis and acute respiratory infections is the cause of resistance to the medications used.

1.3.3 Are there cost-effective means to cure, control or prevent lung diseases?

Most lung diseases are caused or exacerbated by environmental exposures. Theoretically, this means that they are preventable. For example:

The quality of the air we breathe can be assured through:

- Tobacco prevention, aiming to prevent both active and passive smoke exposure;
- Control of ambient air pollution caused by the burning of fossil fuels and vehicle exhausts;
- Reduction of exposure to noxious substances in the workplace.

Vaccine-based prevention is available for some infections and standardised case management recommendations are available for many lung diseases.

- Vaccination against tuberculosis is the most widely administered form of vaccination in the world;
- Effective vaccines against the most frequent causes of bacterial pneumonia are now or will shortly become available;
- Standardised case management for tuberculosis is among the most cost-effective of any health intervention in low-income countries;

- Standardised case management for pneumonia in children under five years of agehas been shown to reduce fatality rates and decrease the inappropriate use of antibiotics;
- Standardised case management for asthma can reduce inappropriate health services utilisation and the cost of care.

All the above knowledge was gained through well-founded research to provide the scientific base necessary for action to improve the health of the community.

CHAPTER 2

EPIDEMIOLOGY, A BASIC SCIENCE FOR LUNG HEALTH

Epidemiology provides the appropriate research methodology and can be considered a basic science for health. To make use of epidemiology and to harness its potential to carry out health research, it is necessary to understand what it is and how it is used.

2.1 What is epidemiology and how is it used?

The essence of epidemiology is that it involves studies of health and disease of groups of people rather than individuals and is concerned with the investigation of the causes and prevention of disease as well as the effective diagnosis and treatment of illness.

It involves counting people and events and classifying them according to common attributes. Its value depends crucially on the accuracy with which this is carried out. This manual is concerned with describing the relevant techniques to do that.

2.1.1 History

Examples of the epidemiological approach to the study of disease and activities to promote health can be found in many historical writings. Hippocrates in his treatise on "Airs Waters and Places" urged his followers to consider the environment in which people live and their personal life-styles in relation to disease. Much ancient written Hebrew law prescribes public health and hygiene action based on epidemiological principles.

In the 17th century in England John Graunt was an early pioneer of the use of quantitative methods in medical science. He used the published Bills of Mortality and Parish Registers of births and deaths in England to calculate age, sex, and cause specific mortality rates for local populations, thus showing the relative importance of different diseases and the impact of contemporary epidemics. After the introduction of routine death certification

in England in 1836, William Farr, the first official medical statistician, analysed the vital data available to him to provide a comprehensive picture of the state of the health of the population and its determinants. This could be applied to the control and prevention of disease.

In the middle of the 19th century, on the basis of his analyses of deaths from cholera in London, John Snow concluded that the likely source was polluted drinking water and recommended effective preventive action. This was long before the agent was isolated.

The principles of epidemiology also apply to assessment of the efficacy of treatment and of preventive interventions. Nowadays no new drug or vaccine is released for general use until it has passed the rigors of a randomised control trial (RCT). The trial of streptomycin in the treatment of tuberculosis published in 1948[6] is often credited with being the first RCT, but there were earlier examples of perhaps less rigorous experiments. For example in 1747 James Lind, a Scots physician, carried out a trial which showed the value of citrus fruits in the prevention and treatment of scurvy among sailors.

In more recent times, over the last 50 years, the concept of epidemiological science has been broadened beyond infectious diseases and their causes and control. It was Sir John Ryle who predicted in 1948 that the "horizons of epidemiology will by degrees be greatly expanded" to include many chronic conditions such as cardiovascular disease, cancers and pulmonary diseases. And so it has been. In addition the same approaches have been applied to the evaluation of health care needs and the effectiveness and efficiency of services provision.

2.1.2 Definitions

Many definitions of epidemiology and of associated technical terms exist which emphasise different aspects of the subject. The following are those given in Last's Dictionary of Epidemiology and are adopted in this text:

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems.

The meanings of terms used in this definition are:

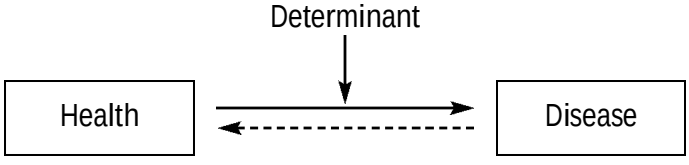
- Study includes surveillance, observation, hypothesis testing, analytic research, and experiments.
- Distribution refers to the analysis by time, place and class of person affected.

(6) MRC Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis. Brit. Med. J. 1948; **2:** 769-83.

- Determinants are all the physical, biological, social, cultural, and behavioural factors that influence health or that bring about change in a health-related state. (Many texts use the word "exposure" in place of "determinant".)
- Health-related states include diseases, behaviours, causes of death, and policy, structure or utilisation of health services.
- Specified populations are those with precisely defined characteristics.
- Application to control indicates the purposes of epidemiological studies – to promote, protect and restore health.

2.1.3 The nature of epidemiological research

Epidemiological research into lung health, as with other diseases, assumes a model of transition between health and disease - or vice versa - which is promoted or retarded by a factor called a determinant. This may represent the development of a disease caused by exposure to a harmful substance or a micro-organism, for example, or it may represent the disappearance (cure) or amelioration (remission) of a disease or its prevention due to a therapeutic or environmental intervention.



This basic diagram underlies the approach to research in epidemiology, the principal purpose of which is to discover more about the determinants of diseases in order to prevent or treat them more effectively and provide optimum care.

The specific aims of epidemiology are to:

- describe the distribution and relative importance of different health-related states and determinants in the population (the descriptive study);
- discover which are the key determinants and to define the natural history of diseases (the analytic study);
- assess the efficacy, effectiveness and efficiency of methods to prevent, cure and alleviate disease (the intervention or experimental study);
- evaluate the process and outcome of services provided for these purposes (health services or operations research).

2.2 Applications of epidemiology

The applications of epidemiology encompass both practical uses related to the planning and evaluation of health services and the better understanding of the causes, management and prevention of disease. The following are examples of such applications:

2.2.1 Measuring the health of a population

Intuitive estimates or impressions of the frequencies of disease and of determinants can be very misleading because of the many factors that may influence such judgements, such as the personal interest of clinicians, patients' access to services and public awareness of health issues. More rigorous methods are needed to:

- measure the burden of disease in defined populations
  Example: in order to plan service requirements and to determine relative priority in resource allocation;
- detect trends in incidence and prevalence of diseases
  Example: differing trends of asthma and of tuberculosis in various populations;
- identify changes in the character of a disease and its consequences
  Example: tracking the emergence of multi-drug resistant tuberculosis;
- define risk groups in the population order to plan appropriate interventions
  Example: characterising environmental agents that cause asthma;
- determine health status, the extent of functional ability or impairment
  Example: measuring the impact of occupational exposures on lung function.

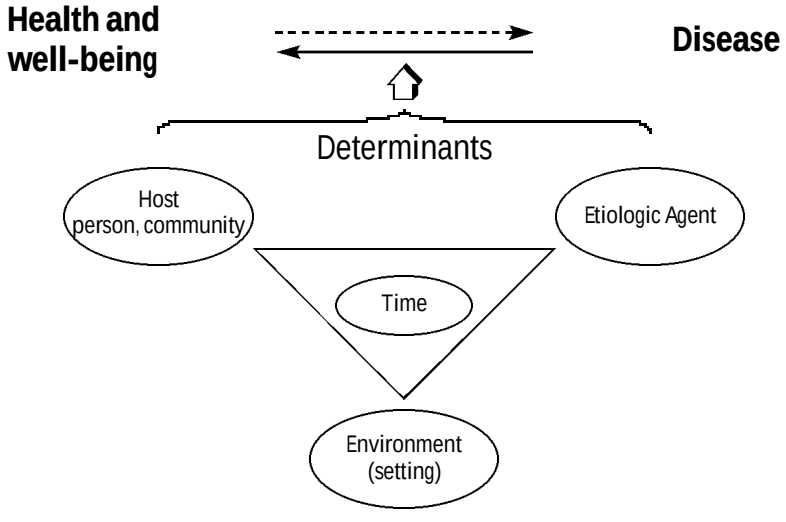
2.2.2 Describing the natural history of disease

The natural history of disease can vary greatly between individuals. Observations on any one person or small number of cases may be unrepresentative. Also the pre-clinical stages of disease cannot be determined reliably in retrospect. Epidemiological studies on unselected groups are required to:

- define a normal range of expected values
  Example: normal values for lung function in a specified population;
- clarify the total clinical picture and identify predisposing conditions
  Example: establishing degrees of severity in acute respiratory infections and in asthma that guide treatment and predict outcome;
- reveal the frequency and course of pre-symptomatic disease
  Example: studies of infection with Mycobacterium tuberculosis in the absence of disease;
- assist in prediction (prognosis) of clinical progress with and without intervention
  Example: clinical trials of preventive chemotherapy in tuberculosis, clinical trials in asthma treatment.

2.2.3 Discovering the determinants of disease

The determinants of disease are often obscure, and more often than not, multiple. Much epidemiological research focuses on identifying possible determinants and establishing the relationship between them and health-related states. Determinants refer to all the physical, biological, social, cultural, and behavioural factors that influence health or that bring about change in a health-related state, as shown in the diagram below.



2.2.4 Controlling and preventing disease

Effective control relies on well-directed intervention in the relationship between the individual and the environment.

This may be achieved in one of four ways:

- remove or destroy the primary agent or determinant. This assumes knowledge of the natural reservoir (or source), mode of spread (or exposure) and site of action
  Example: role of tuberculosis treatment in reducing transmission of infection; substitution of isocyanate-containing paint with other types to prevent asthma;
- protect persons from exposure to the agent by environmental control (hygiene)
  Example: control of nosocomial transmission of infection, reduction of occupational dust and fume exposure by improved ventilation;
- enhance host resistance (protective mechanisms, immunity)
  Example: good nutrition, passive or active immunisation;
- modify human behaviour to avoid risks or promote healthy actions
  Example: prevent tobacco smoking, use respiratory protective devices when unavoidably exposed to harmful agents.

In selecting methods of control or prevention, choose:

- groups shown in descriptive studies to be at increased risk;
- factors which analytic studies show to be quantitatively important;
- methods which experimental studies show to be effective.

2.2.5 Planning and evaluating health services

In the past, services were planned and resources allocated on the basis of historical demand and utilisation, moderated by "judgement" and special pleading. Logical planning and effective administration of health services depends on:

- estimation of need and demand through accurate profiles of morbidity and mortality in the community being served;
- identification of major avoidable hazards to health;
- determination of supply of technical and financial resources;
- accurate measurement of utilisation of health services;
- knowledge of efficacy of interventions;
- evaluation of effectiveness and efficiency (outcomes) of services provided.

The information necessary to implement effective action to improve health (the extent anddistribution of disease, its causes, effectiveness of interventions) is obtained through epidemiological research.

2.3 Understanding association and causation

The results of observational epidemiological studies sometimes demonstrate an association between a suspected determinant and a specific disease or health-related state. This association may even be statistically "significant" (i.e. meeting criteria suggesting that the association is not likely to be explained by chance alone). For example, the results of a study may show a statistically significant difference in pneumonia mortality rates between two different geographic regions.

The fact that such an association was seen is not, by itself, proof of a causal relationship attributable to some specific determinant within the geographic region. The difference may be due to some error or bias in the way the study was carried out.

Ideally proof of causation would be sought through experiments, but this is not usually possible in human populations for ethical reasons. Therefore, when examining conclusions drawn from epidemiologic studies, it is usually necessary to consider evidence from more than one study to evaluate whether or not the associations seen in any one study are causal.

There are various types of evidence that can be used to aid in evaluating a potential causal association. A list of "criteria" for evaluating causation were proposed by Austin Bradford Hill[7] and subsequently developed further. Although no one criterion in isolation may be sufficiently conclusive, several of them taken together can add up to a convincing case.

The following are the criteria listed in the Dictionary of Epidemiology:

Consistency

The association between a disease and a suspected agent or determinant is seen consistently when the results of observational studies are replicated in different settings using different methods. In other words, more than one study has shown similar associations.

Strength of association

This refers to the absolute size of the correlation (or difference between groups) as measured, for example, by the risk ratio. If the disease (or health-related state) is very much more common when the determinant is present the association is said to be "strong". If the disease is only slightly more common when the determinant is present, the association is said to be "weaker". This should not be confused with "statistical significance". Both weak and strong associations may be statistically significant.

Specificity

This is established when a suspected agent or determinant (or combination of more than one determinant) consistently produces the same specific effect.

Dose response

An increasing level of exposure to the agent or determinant (in amount and/or duration) increases the risk of developing disease.

Temporal relationship

Exposure to the determining factor always precedes the outcome. This is an essential criterion though the date of (initial) exposure can sometimes be hard to ascertain.

Biological plausibility

The association is consistent with current understanding of pathobiological processes related to a suspected agent or determinant. This criterion should be applied with caution since the relevant process may not be currently understood.

(7) Hill AB. The environment and disease: association or causation. Proc. R. Soc. Med. 1965; **58:** 295-300.

Coherence

The association should be compatible with existing knowledge of the natural history of the disease.

Experiment

The condition can be prevented or ameliorated by an appropriate experimental intervention (e.g. removal of the suspected agent or determinant from the environment).

See the companion text for a practical exercise on cause v. association.

CHAPTER 3

GETTING STARTED IN RESEARCH

3.1 The research protocol

3.1.1 Why focus on the research protocol?

In laboratory research, scientists generally employ an experimental design with the following steps:

- propose a hypothesis;
- write down methods or procedures;
- carefully and consistently apply the procedures to the experimental group or series;
- study a control group or series subjected to the same conditions as the experimental group, minus the specific study procedures;
- compare the results (outcome) between the experimental and control groups.

Conclusions from laboratory investigations following this approach are usually considered to be more "scientifically sound" than conclusions derived from simple, uncontrolled observations. In the following section, the elements of a typical research protocol are introduced. These will be explored in more detail in the subsequent sections of the course notes.

3.1.2 Typical format and elements of the Research Protocol

In population based research, despite the fact that many of the studies are observational rather than experimental, it remains essential to elaborate, and follow, a research protocol for the same reasons as in laboratory research. Doing so will increase the likelihood that the conclusions drawn from the research will be scientifically sound.

1. Abstract

2. Study description

- a) Study Question
- b) Rationale, previous studies on the subject
- c) Aims and Objectives
- d) Design and Methods
  - Study design
  - Study populations
  - Sample size and statistical power
  - Subjects: selection and definitions
  - Data collection methods: measurements, definitions
  - Data management and statistical analyses
- e) Project Management
  - Personnel required
  - Duration of the study (timeline)
  - Follow-up procedures (if needed)
- f) Strengths and limitations
- f) References

3. Ethical Considerations

4. Significance (or expected impact)

5. Budget

6. Investigators: role of each and curriculum vitae

This course is designed to assist you in developing, writing, (and ultimately following) a scientifically sound research protocol in the domain of population based research.

The research protocol can be totally fluid during its development phase but, once agreed, it must be strictly followed.

Abstract

This should be concise but sufficient to orient the reader to the main purpose of the study, how it will be conducted and its expected benefits. It is, as it were, a sketch plan of the study that will help the assessor see the general plan before examining the details. It is placed at the head of the protocol, but is often written after the protocol itself is completed.

Study description

The description of the study should provide all the information necessary for those assessing its merits to address the following questions:

- What question do you hope to answer, and what is your expectation (or hypothesis) about the answer to this question? Is this research question (and hypothesis) important? How does it fit with current knowledge in the field?
- Will the study, as described in the protocol, answer the research question?
- Can the particular objectives be met with the study design and methods proposed in the protocol?
- Can the study actually be carried out with the resources identified?
- When the study is completed and results are being analysed, will there be sufficient statistical power to make valid conclusions about the study question?
- Will the findings impact favourably on health, either directly or indirectly?

It is useful to keep the above questions in mind throughout the process of preparing your protocol. Put yourself in the position of the reviewer!

The essential elements of a protocol as outlined above should explain the study in terms of answers to the following questions:

- WHY? Sets out the study question and the relevant background information;
- HOW? Describes the study design and the rationale for choosing it;
- WHO? Defines the target and study populations and sample size;
- WHAT? Identifies the variables to be measured, instruments to use and outcomes to be analysed;
- SO WHAT? Comments on the expected significance of results and contribution to knowledge.

Study Question

Your protocol should start with a clear and precise formulation of the research question. It is good practice to write this in the form of a question not a statement.

Example:

Why is asthma among children in Istanbul exceptionally frequent?

Rationale, previous studies on the subject

The purpose of this section is to state how the research question arose from current knowledge about the subject. The progression of your ideas needs to be set out in a logical sequence. Be concise; include key references, not a complete review of the literature.

- Discuss the importance of the topic;

- Review the relevant literature and current knowledge (including deficiencies in knowledge that make the study worth doing);
- Describe any results you have already obtained in the area of the proposed study;
- Indicate how the research question has emerged and fits logically with the above;
- Outline in broad terms how you intend to address the research question;
- Explain how your study will add to knowledge and help to improve health and/or save money.

Objectives, Hypotheses, and Specific Aims

Objectives

Even a precise study question is often too broad for one study to answer. For example, if your study question were "Why is asthma among children in Istanbul exceptionally frequent?" you could not hope to study all possible answers to this question in one study. Therefore, you must break down the question into one or more objectives for your particular study.

Example:

> The objectives of this study are to determine if the excess asthma in Istanbul is related to a combination of genetic predisposition (estimated by atopic status) and socio-economic status and / or indoor air pollution.

Hypotheses

To meet these objectives, the study plan may include comparisons of disease or exposure rates between more than one group of subjects, using statistical testing to evaluate the comparisons. If this is the case, the objectives should also be stated in the form of hypotheses to be evaluated by the statistical tests. The hypotheses should be written as statements to be refuted. These are referred to as null hypotheses.

Example:

> The following hypotheses will be tested:
> 1. Asthma prevalence rates are not different among children from low and high socio-economic groups in Istanbul.
> 2. Asthma prevalence rates are not increased in children living in homes with increased air pollution.
> 3. The relationships between asthma prevalence rate and socio-economic status and between asthma prevalence rate and indoor air pollution do not differ according to the atopic status of the child.

Specific Aims

Following this, you summarise the practical steps the study team will need to carry out toaddress the objective of your proposed study. These are often called Specific Aims.

Example:

To meet these objectives, the following specific aims have been identified:

1. Identify a suitable source of childhood asthma cases and select 200 cases, following a specified case definition.
2. Identify and select suitable control subjects (individuals without asthma).
3. Record personal, demographic, and socio-economic information about cases and controls using a standard questionnaire.
4. Perform allergy skin tests on cases and controls (as an indicator of atopy).
5. Measure indoor particulate exposure on each of 3 randomly selected days for each participant.
6. Compare risk ratios for atopy, low socio-economic status, and increased indoor air pollution between cases and controls.

It should be obvious that you cannot complete this section of your specific aims immediately. You will have to decide on your study design and methods first, and then list your specific aims later. They are placed here to make it easier for the reviewer of your protocol to quickly and easily understand your study plan.

Design and methods

Study design

This should state the selected design of the study. Keep in mind that the study design is chosen in relation to the study objectives. Explain why the particular study design has been chosen in preference to other possible designs.

Study population

This section outlines the setting for which the research has relevance. When the results are obtained, to whom do they refer? Which groups have the kind of disease studied? Are all communities affected? Is it only one geographical, professional, age group? This section also describes how one can be certain that the results can be generalised to the population identified.

Example:

- Which children in Istanbul should we study? All children? One school? Asthmatics who are seen in hospitals only? Any child who reports symptoms which are consistent with asthma?
- Will the results from our study population be able to tell us something about all children? About all children in Turkey?
- Who are "children"? Those under 16 years of age? Under 2? Are there a lower and an upper limit?

Sample size and statistical power

It is necessary to estimate how big the study needs to be to answer the question posed. Specify the assumptions made for the calculation, and include a table of the calculation of sample size (and power) given varying assumptions.

Example:

- How many children will be recruited?
- What proportion of the whole population do they represent?
- Will this number of children be sufficient to answer the question?
- What is the variation in the prevalence of the factor (air pollution, atopy) among the children?
- How big a difference in children with/without the factor can I detect?
- Is this an "important" difference?

Subjects: selection and definitions

This section should provide:

- a detailed explanation of how many subjects in which categories will be recruited into the study, where and why;
- definitions of eligibility, of inclusion and exclusion criteria, criteria for discontinuation;
- realistic estimates of the numbers of potentially eligible subjects;
- description of mechanism of recruitment;
- discussion of the feasibility of recruiting the required number of subjects and realistic estimates of the proportion that will agree to participate.

Example:

- Will children with (without) asthma be included?
- Which children will be excluded?
- How will the children be approached?
- What will you do if they decline?
- How many children will be eligible and what percent do you expect will actually participate?

Data collection methods: definitions and measurement

It is essential to state how the data will be collected to determine both the health outcomes (disease or other "health-related state") and the determinants you are planning to study. This means specifying exactly how these will be measured or defined in the proposed study. Quality control procedures should also be specified. If the procedure is a standard one that has been described before, it should be referenced.

This includes:

- precise definitions of all terms;
- consideration of pilot testing for methods and instruments;
- discussing the validity and reliability of the definitions proposed;
- discussing the limitations of the measurement tools and definitions proposed (e.g. the effects of error or misclassification).

Example:

- How will asthma be determined for this study: by questionnaire, clinical examination, reviewing patient charts, hospital admissions?
- What will happen if someone who doesn't have asthma is said to have it?
- What is the definition of indoor air pollution for this study? How will it bemeasured or estimated?
- Which components of indoor air pollution will be evaluated?
- How will socio-economic status be determined?
- What will be the definition of atopy for this study?

Data: management and statistical analysis

This section will describe:

- procedures for coding and entering data into computer files;
- measures to ensure the completeness and accuracy of the information;
- examples of how the results will be displayed and comparisons made;
- tests to be used to carry out statistical analyses in order to test each of the stated hypotheses.

It is useful to structure this section according to the research questions, objectives, and hypotheses to be addressed.

Include the appropriate reference for the statistical tests: the statistics book or article in which the method is described, or the statistical computer program to be used.

Example:

- How will the patient records (or questionnaire responses) be abstracted?
- How will you ensure that the questionnaire data are correctly coded?
- What types of tables, graphs and figures will be used to display the results?
- What statistical tests will be used to test the hypotheses of the study?

Overall Project Management

This section will identify the

Personnel required to carry out the research and define their tasks. It will justify the personnel proposed in terms of the tasks and the amount of time required. It will specify the responsibilities of each staff member.

Duration or timeline will set out the anticipated time required for each phase of the study including:

- pilot testing;
- recruiting of subjects;
- preparation of forms and questionnaires;
- data collection;
- follow-up procedures;
- data checking and statistical analyses;
- reporting – to participants, to sponsors, to the community involved, to the academic community.

Follow-up procedures for study participants should be specified where appropriate (e.g. will each study participant be informed of the study results? How will you respond if you uncover clinically relevant disease in a participant who is not being treated?).

Strengths and limitations

In developing the protocol, many compromises will have to be made in choosing among several possible study designs and approaches to collecting information. There is no such thing as a perfect protocol and there will always be aspects of the protocol that are open to criticism. It is important to include a section in which you address the possible criticisms of your design and methods and provide reasons why you think the limitations imposed by your choices are not serious ones. Similarly, it is useful to identify those aspects of the specific protocol that you think are particularly strong and worthy of financial support.

References

This section will list references from all the sections discussed above (including rationale, previous research, standard methods, statistical analyses) normally listed according to the order of their presentation in the protocol.

Ethical considerations

This section must explicitly state that the principles of the Helsinki Declaration have been taken into account and will be followed. It indicates:

- how the quality of the technical aspects have been assured;
- the expected hazards of the study procedures and their expected benefits;

- the rationale and justification for carrying out the research;
- the priority of the participants' interest over those of science or of society;
- how these interests will be safeguarded;
- responsibility for liability for injury to study participants;
- how the participants are informed of the study and;
- how they give voluntary consent to participate.

Example:

- Will any child be at risk from a reaction to the allergy skin tests?
- Who will provide consent for the child to participate?
- Who will know about the results of the examinations?
- What will be done if disease is found?

If the research is sponsored from outside the community or the country, it must explicitly outline the interests of all parties in the research and the benefit to participants, to the local investigators and to the community.

- To whom do the study results "belong"?
- Who will be the authors of scientific publications?
- What will the sponsors do with the results?
- Will ill people who are diagnosed in the study get proper care?

Significance (expected impact)

This section restates the justification for the study in terms of the anticipated results. It will specify:

- the implications of the potential results;
- how the results of the study may be used by your own research team in the future, by other researchers, by policy makers, by the community.

Budget

Each item of expenditure expected in the conduct of the study must be specified, even if the cost is covered by routine operations of the health service or by other sources outside the study itself. Expenditures should, as much as possible, be given in units (such as salaries per hour according to qualifications).

A written budget justification may be included to explain various expenditures in further detail.

Investigators: role of each and curriculum vitae

This section should describe what role each investigator plays in the study and should state clearly who is responsible for each component of the study.

The curriculum vitae should provide a clear description of the qualifications and experience of the investigators, including training, academic degrees or certificates, research experience and scientific publications.

This section should also describe the other responsibilities of the investigators. This is to assure assessors that the investigators are sufficiently experienced to carry out the research and to make it clear that the investigators have the time to devote to the work to complete it.

3.2 Deciding what to study: The research question

Choosing a research question is an essential first step in research. Choosing the right question is necessary to clearly outline your protocol and will determine the likelihood that you will get someone to give you money to pay for the research, an indispensable requirement in doing research.

3.2.1 Basic concepts in framing a research question

The topic of research is traditionally formulated into a question. Why is it done in this way?

The question

The goal of research is to increase knowledge. The description of the specific knowledge being sought and the methods used must be precise in order to formulate an approach to creating that knowledge. The precise description is what we mean by a research protocol.

The topic of the research is formulated into a specific question that must be clearly defined as specific objectives. The more precise the question, the more likely it is that research will provide new knowledge. By formulating the topic into a question, it is easier to outline the steps necessary to arrive at an answer to the question. If the topic is stated but not precisely defined, it is less clear how the knowledge will be derived.

Example:

Let us say we are interested in multi-drug resistant tuberculosis in Bangkok. We could outline the research topic as:

Multi-drug resistant tuberculosis in Bangkok

Or

Is multi-drug resistant tuberculosis in Bangkok caused by inadequate case management?

The advantage of the question over a simple declaration of the topic area is obvious. The question format requires more precision and leads to a logical approach to the research topic.

The hypothesis

If the answer to the research question will involve comparisons between two or more groups, or correlation among factors, the statement of one or more hypotheses must follow the formulation of the research question and the study objectives. The purpose of a hypothesis is to set the research question into a format amenable to answering through a set of statistical tests.

The hypothesis is posed as a statement that involves a comparison or an association, rather than as a question. It is traditionally stated in the form you anticipate wanting to refute or reject. This form is called the null hypothesis. The reason for this is that the result provided by most statistical tests is the probability of erring if you reject the null hypothesis. Rejecting the null hypothesis increases our confidence, with a given level of probability, that there is a relation between the variables (health-related states, agents, determinants) studied.

Example:

Research question – Is multi-drug resistant tuberculosis in Bangkok caused by inadequate case management?

Research objective – To compare the rate of multi-drug resistant tuberculosis in a group of patients in Bangkok who were given inadequate case management with a group who received adequate case management.

Research hypothesis – The prevalence of multi-drug resistant tuberculosis is not different between those who were given adequate and those who were given inadequate case management in Bangkok.

The progressive clarity achieved by formulating a topic into a question, a question into more precise objectives, and then formulating the hypotheses should be obvious. The question sets the framework, the objectives force us to be precise and to propose key terms such as multi drug resistant, tuberculosis, case management, quality of case management, Bangkok.

Posing the research hypothesis forces us to think carefully about what comparisons will be needed to answer the research question, and establishes the format within which we willapply our statistical tests when interpreting our results. In this instance, we set about to establish the association (or lack of it) between inadequate case management (the

"determinant") and multi drug resistant tuberculosis (the "health-related state"), by comparing rates in groups of patients with different case management approaches. In determining the nature of the association between them, we can evaluate the strength of the association (i.e. how large a difference there is in the rates of multi-drug resistant tuberculosis) and estimate a level of confidence that the difference found in our study is true. These are two key elements in considering causation.

3.2.2 Points to consider when choosing a research question

In some senses, science may be considered a purely "intellectual" activity and the pursuit of knowledge should be valued for itself. We do not, however, live in a world of unlimited resources and so priorities must be established. Even those resources that exist are often distributed through obscure mechanisms and networks. When selecting a research question, therefore, we need to understand prevailing operational realities and how to work within their limits.

Points to consider when selecting a research question:

- Relevance to health in the general community (frequency, seriousness);
- Likelihood of effective intervention or implementation of the findings;
- Feasibility of successfully carrying out a study to answer the question;
- Generalisability of the findings to other settings;
- Interests and experience of the research team ;
- Relevance to political and social imperatives (community, scientific).

See the companion text for a practical exercise on choosing research questions for protocol development in the course.

CHAPTER 4

STRUCTURING RESEARCH: STUDY DESIGN

4.1 Study design: its relation to the development of medical knowledge

The aim of research is to gain knowledge and understanding of the natural world and how it functions. This requires the assembly of systematic observations into logical patterns that permit comparisons. Such comparisons allow us to draw conclusions that explain observed realities (deductive reasoning) and to formulate ideas that explain general realities (inductive reasoning). We then test these ideas by applying them to other experiences.

The process of development of such concepts in medical science has traditionally followed a series of systematic steps. It often begins with observations on a person with the disease (the case). Assembling a number of cases of the same disease (the case series) then identifies the characteristics specific to a disease.

Recording information about the features and natural history of a disease or condition in a larger population, using a standard approach, provides a more thorough picture of the burden of the disease in the community (the descriptive study). This can provide insights into possible determinants of the disease (hypothesis generation). Examining these possible determinants by carrying out studies involving planned comparisons between locations or within subgroups in a given population provides more convincing evidence (the analytic study).

Finally, we move from observing patterns and making comparisons among groups to testing the effects of an intervention on the characteristics of a group and comparing them with another group to which the intervention has not been applied (the experimental study).

Experimental studies

These studies are most like the classical laboratory studies performed by researchers in other fields. The difference is that human populations are used. The use of this design is limited primarily by ethical constraints. Examples of these are clinical or intervention studies.

Observational studies

Since it is frequently neither feasible nor ethical to use humans in experiments, most epidemiological research involves observing (not experimenting on) human populations.

The goal of observational studies is "to observe and measure, systematically, the results of natural experiments". The challenge is to find a study design that will allow the research team to make such observations.

Descriptive studies

These studies are concerned with describing existing conditions, usually with the aid of data already collected for other purposes (e.g. death or disease registry data, infectious disease records, hospital admission data) without specific hypotheses or planned comparisons formulated in advance.

Analytic studies

In these studies the objective is to test hypotheses, using comparisons between two or more groups, or associations among various factors in one group.

Cross-sectional studies

- most like descriptive studies;
- data on both disease and determinants are collected for a specified population at the same point in time;
- comparisons are made between subgroups to test specified hypotheses.

Cohort studies

- a population is identified and divided into two groups, those with and those without a specified determinant (risk factor, exposure, etc);
- the two sub-populations (with and without determinant) are followed over time and disease rates in each sub-group determined and compared to test hypotheses.

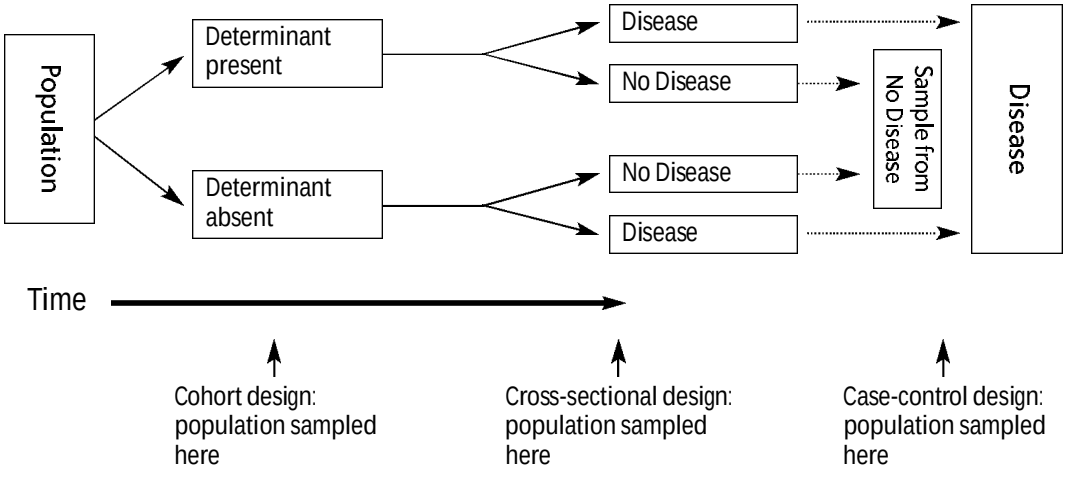
Case-control studies

- persons with a given disease are identified (cases);
- the population from which the cases arose is identified and a sample, or subset, of this population (without the disease) is selected (controls);

- the frequency of various determinants is measured in the cases and controls (and compared to test specified hypotheses).

See the companion text for a practical exercise on identifying study designs.

4.2 Overview of study architecture



The basic framework of epidemiological study designs is that in which the presence or absence of the health-related state (usually disease) is related to the presence or absence of the determinant (e.g. treatment, exposure, personal risk factor) under investigation.

From these two components (disease and determinants) four categories can be defined:

1. determinant (or risk factor, exposure, treatment) present, with disease,
2. determinant (or risk factor, exposure, treatment) present, without disease,
3. determinant (or risk factor, exposure, treatment) absent, with disease,
4. determinant (or risk factor, exposure, treatment) absent, without disease.

4.3 Study design details

4.3.1 Descriptive studies

Descriptive studiesprovide a useful initial overview of a problem. They can help to give a valuable perspective on the burden of disease and may assists in planning of health care and preventive services.

Aims

- to measure the relative importance of different health problems in a community and monitor changes;
- to describe the frequency of different diseases or health problems within a community.

Study Methods

Descriptive studies assemble morbidity or mortality statistics, usually from routine sources, or data on health-related variables (e.g. lung function, health services utilisation) and report the information according to meaningful categories, such as:

- Time (when it occurs)
- Place (where it is found)
- Person (who is affected)

Descriptive data is usually grouped under these headings, looking for variations that correspond to patterns in the prevalence of possible determinants.

Time

The incidence of diseases may vary on several different time scales:

- Secular trends - changes over many years; e.g., tuberculosis, asthma;
- Cyclic changes- rises and falls at regular intervals; e.g., whooping cough, measles;
- Seasonal variation - increases at particular times of the year; e.g. asthma, influenza;
- Epidemics - irregular temporary increases; e.g. legionnaires' disease.

Place

The spatial aggregation of cases of a disease may occur in several dimensions:

- Geographical - variation between countries and regions with different climates, ethnic groups, social, cultural and dietary customs; e.g., lung cancer, asthma;
- Urban-rural - differences in crowding, poverty, environmental pollution, diet and occupation; e.g., tuberculosis, chronic obstructive airways disease;
- Locality - housing, environmental pollution, types of domestic heating and cooking fuels and adequacy of ventilation which may differ between localities and households in the same community; e.g., pneumonia, airways obstruction;
- Institution - disease incidence may vary within institutions, such as schools, military camps, prisons, factories and hospitals according to organisational units such as classes, barracks, wards and workshops; e.g., acute respiratory infections, occupational lung diseases.

Person

Various personal characteristics may influence the risk of contracting disease by affecting either an individual's chance of exposure to causative agents or the individual's susceptibility:

- Age and gender - physiological and anatomical differences, immunological capacity, nature and duration of exposure to agents, past disease and disability, degenerative processes; e.g., tuberculosis, pneumonia, asthma;
- Marital status - life style, contact with children; e.g., acute respiratory infections;
- Ethnic group - genetic susceptibility, environment, cultural practices; e.g., asthma, tuberculosis;
- Family - shared genetic inheritance and environment, bonds and traditions e.g., asthma, tobacco smoking;
- Occupation and socio-economic status - exposure to noxious substances and infectious agents, place of residence, travel, economic resources; e.g., asthma, pneumonia, chronic dust-related lung disease.

Advantages

- cheap and quick, cost-effective use of existing information;
- useful initial overview of a problem;
- identify parameters for further study.

Disadvantages

- Difficult to identify all cases, especially those that are rarely fatal or not usually medically managed;
- Data on disease and related variables may not be available or not in required form;
- Methods of data collection and diagnostic criteria not standardised.

4.2.2 Analytic studies

The analysis of descriptive data often leads to hypothesis formation based on observed correlations between diseases or health related states and possible determinants.

The hypotheses generated need to be tested by planned field studies designed to examine if:

- people with disease have common attributes, or have been exposed to the same agent;
- people with particular attributes who are exposed to a specific agent are more likely to develop disease than those who do not have those attributes.

This is the role of analytic studies of which there are three types:

- Cross-sectional studies;
- Cohort studies;
- Case-control studies.

Cross-sectional Studies

Definition

A study in which all the measurements are taken at a particular point in time. These measurements identify cases (or other health-related states) as well as suspected determinants (past or present exposures, or specified attributes) and compare subgroups in the population.

Aims

- To test hypotheses on disease causation by showing the degree of correlation between a range of possible determinants (e.g. personal attributes, exposures, behaviours, treatments) and the presence of disease either:
  - by comparing disease prevalence in groups with different types or levels of the determinant within the study population,

    Or

  - by comparing disease prevalence in a population with the determinant with that from data obtained in a similar population without the determinant (or with "expected values" based on regional rates).
- To assist health service planning by providing systematic, comparative, information on
  - the burden of disease and disability in various subgroups in the population;
  - the groups within the population at greatest risk on whom treatment and preventive services need to be focused.

Study Methods

The purpose of the study must be clearly defined, as it will critically affect the choice of study population and methods.

The research protocol should:

- define the target population and how the subjects for study will be sampled (see section on "choosing study populations" below);
- describe how the various subgroups for comparison will be identified (e.g. disease absent/present; exposure absent/present) with specific criteria for inclusion and exclusion of subjects;
- provide clear definitions for how disease and exposure will be measured and how subjects will be assigned to the various disease and exposure categories (operational definitions);

- prescribe the study instruments and the training of the research team in their use and study procedures;
- set out plans for data management and analysis, including error checks and other quality control procedures.

Advantages

- results can be obtained relatively quickly and cheaply;
- large numbers of possible associations can be explored;
- standard methods of measurement of both exposures and outcomes can be used.

Disadvantages

- temporal relationships (between potential determinants and disease) is not always clear;
- when disease is relatively rare a large study population is required;
- recall of past events may be unreliable;
- population being studied comprises survivors of a cohort who may be a biased sample of the target population.

Cohort Studies

Definition

A cohort is a group of persons who share a common experience. A cohort study is one in which a population (cohort) of persons who are free of disease is defined at a point in time, and subsequently classified according to the presence or absence of a determinant (e.g. exposure to an agent). They are then observed over a period of time to identify the subsequent appearance (incidence) of disease in those with and without the determinant (exposure).

Aims

- to show whether those with a particular determinant (exposed to a specific agent or with a particular attribute), have a greater risk of subsequently developing disease as compared with those without the determinant (non-exposed persons or those without the relevant attribute), ie. to identify risk factors associated with increased or decreased disease incidence;
- to measure the excess risk (risk or rate ratio) that may be attributable to the determinant.

Study Methods

The cohort may comprise a sample of one or other of the following:

- the general population;
- a group known to have a high incidence of the disease under study (e.g. those in a particular age, sex or occupational group);

- persons with special attributes which facilitate their identification and ease of follow-up (e.g. doctors, nurses, or other employees in industries with a nominal roll and low turnover).

If the study plan calls for the cohort to be identified and recruited "today" and followed into the future, the study is called a prospective cohort study.

All participants are interviewed at the outset using standard methods (e.g. questionnaire, standardized tests) to record personal attributes, past histories, and relevant biological variables. Determinants (such as exposure to suspected harmful agents) are recorded either continuously or at specified points in time during the follow-up period.

Controls in prospective cohort studies are often "internal", i.e. members of the cohort who are not exposed to the determinant under investigation. Where the cohort comprises a group of individuals all of whom are or have been exposed (e.g. miners and coal dust) an "external" comparison group is required. This may be the total population of the region or a non-exposed population with otherwise similar attributes.

After a period of time has elapsed, the disease rate or average disease measure is calculated for each group, and compared.

If the study plan calls for the cohort to be selected based on their membership in the population at a specified point in the past, and followed from that point (in the past) up to "today", then the study is called a retrospective cohort study.

The study population is enumerated based on membership in the specified population at a point of time in the past (using existing records – i.e. employment records, school registries, town lists).

Information to classify each study member as exposed or not (to the agent being studied) is usually collected from records, but may be collected from the subject him or herself or from a family member.

For studies of non-fatal outcomes, study subjects are located today, and their disease status is determined. For mortality studies, the vital status of each study subject is determined today (using public records) and for those who have died, the cause of death is determined.

The incidence rate of disease (or death) is calculated in each group and compared, often by examining the ratio of incidence rate in two groups (relative risk). This is discussed in greater detail below.

Advantages

- proper temporal sequence of events can be observed, thus helping to distinguish causes from associated factors;

- new cases of disease can be identified in a specified population and time period, which allows incidence to be calculated;
- several possible determinants and outcomes can be studied simultaneously;
- determinants and outcomes can be measured precisely.

Disadvantages

- large population required if incidence is low;
- long time-scale before results emerge especially if incubation period is prolonged;
- relatively expensive in resources;
- losses from population during study may bias results;
- standard methods and criteria may drift over prolonged follow up.

Example

Study objective:

To investigate the role of exposure to wood smoke in the development of asthma in children.

Prospective study:

- identify all children without asthma in all 50 nursery schools in our town, as of today's date,
- divide them into 2 groups:
  - those who live in houses where wood is used for fuel(exposed),
  - those who live in houses where wood is not used for fuel (unexposed),
- follow them for 10 years, checking up on them each year,
- after 10 years, calculate the incidence rate for the development of asthma in each group and compare these (risk ratio or risk difference).

Retrospective study:

- using school and public health records, identify all children who were attending any one of the 50 nursery schools in our town, 10 years ago,
- using address information from the records, divide the children into two groups:
  - those who probably lived in a house where wood was used for fuel (exposed)
  - those who probably lived in a house where wood was not used for fuel (unexposed),
- locate the children today (e.g. at their current school), and identify which children developed asthma during the past 10 years, and which did not,
- calculate the incidence rate for asthma in each group, and compare these (risk ratio or risk difference).

Case-control Studies

Definition

A study in which the frequency of a determinant (history of past exposure to a possible agent and / or personal attributes) in a group of persons with a disease (cases) is compared to the expected frequency of the determinant (exposure or attribute) in the population that gave rise to the cases. The "expected frequency" is usually determined by studying a group of persons without the disease (controls).

This study design is often used as a method of preliminary investigation of a hypothesis because of its significant advantages over other study methods (see below).

Aims

- to show whether the agent or determinant is found more frequently among those with the disease than among those without;
- to estimate the relative risk of any excess frequency using the "odds ratio".

Study Methods

The essence of a case-control study is to obtain unbiased data from representative cases and controls drawn from the same population in order that fair comparisons of their exposures to risk can be made.

Data on the histories of past exposures and relevant personal attributes may be obtained by direct questioning or from records.

To avoid bias it is essential to elicit data and make observations on controls in exactly the same way as for cases.

Selection of Cases

- Ideally select all cases in a defined population though it is usually only practical to recruit a sample of cases from an easily accessible source;
- No matter how they are chosen, it is essential that all persons with the disease in that population have an equal chance of being identified and selected.

Commonly used sources of cases include

- patients attending hospital or family doctor;
- patients listed in a register (e.g. cancer register);
- persons identified by systematic screening for the disease in a defined community;
- death certificates.

Selection of Controls

- because the purpose of the controls is to give an estimate of the "expected frequency" of the determinant, controls should be **a representative (ideally random) sample of the population from which cases were recruited;**
- after selection, controls must not be discarded or replaced unless erroneously included;
- similarity in other relevant attributes can be ensured by "matching" on potentially confounding variables, e.g. age, gender, social class, marital status. (Note that the determinant role of all matched variables and any variables correlated with them cannot be assessed in the analysis. There is a danger of "over-matching" where risk factors and potential confounding variables are not known. When in doubt it is better not to match and adjust for confounding in analysis);
- statistical power can be increased by use of 2 or 3 controls per case.

Frequently used sources of controls include:

- persons living in the same locality or from the same work place, pupils attending the same school etc.;
- population registers, e.g. birth, school and electoral rolls, family doctor lists;
- hospital patients (where cases are drawn from hospital attendees) who have other unrelated conditions;
- relatives - accessible but have similar environmental exposures and genetic profile which may confound comparisons;
- random digit telephone dialling.

Advantages

- results can often be obtained more quickly and cheaper than with cohort studies;
- the size of population required is economical;
- often easy to identify a relevant case group;
- the only practical method for study of rare diseases.

Disadvantages

- temporaal sequence of events not always clear;
- cannot measure incidence or prevalence rates of disease as the total population denominator is not known;
- difficult to ensure that controls are representative of the population giving rise to the cases (confounding);
- incompleteness of records and unreliability of recall of past events and past exposures, especially in controls.

Example

Study objective:

- to investigate the role of exposure to wood smoke in the development of asthma in children;
- identify all the children (aged 6-10 years) with asthma who have been treated in the past year in the major health clinics in our region (cases);
- identify a comparison group of children of the same age, who do not have asthma and who attend school in the same region and select a random sample of them;
- interview the parents of each group of children to ask if they used wood for fuel at home when the child was born and during the years the child was between age 0-6 years.

Calculate the frequency of wood fuel use in the homes of each group and compare these.

Analysis

Because the size of the underlying population is not recorded in a case-control study, it is not possible to calculate the incidence or prevalence rates of disease in a case-control study. Therefore, we calculate and compare the frequency of exposure in the disease and not-disease groups, using the odds ratio. This is described in the section on "indices of risk" below.

Where more than one factor is involved as a cause or determinant, or where there is interaction between them, more complex statistical tests are useful to evaluate their relative etiological importance (e.g. logistic regression).

"Nested" case-control study

If a study population has been enumerated for a retrospective cohort study, but there are limited resources and it is not possible to obtain detailed information on potential determinants (exposures and personal attributes) for the entire cohort, a nested case-control study can be carried out.

This is a case-control study, in which the cases are all the persons from the cohort who have developed disease, and controls are a random selection of the non-cases from the same cohort population. The collection of detailed determinant information is then confined to these individuals rather than the entire cohort.

It takes advantage of the cohort study approach by defining exposure and other risk factors or determinants at the outset of the study.

Advantages

- no exposure recall bias;
- reliable pre-illness data;
- cheaper than cohort study.

These to some extent offset the disadvantages of normal case-control studies listed above.

4.2.3 Experimental (intervention) studies

The results of case-control and cohort studies may suggest possible causes of disease andmethods of disease prevention and treatment. The efficacy and safety of relevant interventions need to be formally tested. Such studies normally take the form of intervention studies or clinical trials.

Definition

A study in which a population is selected for a planned trial of an intervention whose effects are measured by comparing the outcome in an experimental group (receiving the intervention or treatment) to that in a control group receiving "conventional" treatment or placebo (no active intervention).

Aims

- to evaluate both the beneficial and adverse effects of medical interventions;
- to examine causal hypotheses.

Most intervention studies are in the form of randomised control trials (RCT). These evaluate specific interventions whether therapeutic or preventive by comparing of outcomes in test and control groups to which individuals in the trial population have been allocated at random.

RCT's are used to

- assess the efficacyand safetyof a new treatment or preventive intervention compared with a control regimen (conventional management or placebo);
- compare alternative treatments;
- evaluate the effectivenessand efficiencyof different forms of service provision;
- provide direct evidence that exposure to a suspected agent causes disease or that its removal prevents or reduces the frequency of disease, but such trials can present ethical problems.

Essential requirements of a clinical trial are that

- outcome benefits should be clinically important and measurable;
- the intervention must be compatible with the health care needs of the patients in the trial;
- there must be reasonable doubt regarding the efficacy of the intervention;
- the intervention must be acceptable to both patient and provider;
- there should be reasonable belief that the benefits will outweigh risks.

Clinical Trial Design

- similar in principle to that of a cohort study;
- the population under study should be representative of the population in which the intervention will be applied (target population). It should be stable and accessible. Volunteers are not acceptable;
- subjects must be allocated at random to test or control groups (sometimes it is desirable to stratify randomisation - see below);
- the intervention is applied to the test, but not to the control group, who may receive placebo or conventional treatment;
- to avoid bias in reporting illness (or other relevant events/behaviour) or in assessing outcomes, ideally neither the subject nor the assessor know to which group the individual participant belongs (double blind). Sometimes this is impractical because of the nature of the treatment (e.g. surgery) in which case a single blind design may be adopted where either the patient or the person who assesses the outcome does not know the treatment;
- the effect of the intervention is assessed in terms of one or more clearly defined primary or secondary outcomes using the same criteria in both treatment and control groups;
- outcomes should always include adverse events as well as beneficial effects.

Misclassification of outcomes in either group will reduce the size of any difference between the two groups and may lead to a spuriously reduced apparent benefit.

Follow up procedures should be identical in test and control groups, as follows:

- outcome definition and methods of data collection should be simple and sufficiently sensitive to detect all relevant events;
- outcome data must be collected and recorded in a standard manner;
- follow up must be equally rigorous in test and control groups;
- follow up starts at allocation and continues long enough to determine the outcome in all subjects;
- all losses to follow up must be reported and every effort made to minimise them.

Analysis

All patients randomised must be included in the analysis - this is called intention to treat analysis. It means that:

- patients are analysed in the group to which they were originally assigned irrespective of any enforced change in their management that departs from the protocol;
- all events throughout follow-up are counted;
- all outcomes specified in the protocol, both beneficial and adverse, are analysed.

Efficacy = (E-I / E) x 100

where E = expected incidence (derived from the control group) and
I = intervention incidence.

Sequential analysis is a method of analysis whereby the comparison of outcomes between test and control groups is continuously monitored. It allows a study to be halted when a statistically significant positive or negative result has been obtained. It is used when a result is required urgently or serious adverse events are possible or anticipated benefits are high. It also allows economy of resources by avoiding an unnecessarily prolonged study.

Ethical considerations

Like all epidemiological studies, clinical trials must be ethically evaluated before being approved. Consideration should be given to:

- unknown nature of possible risks of treatment with a new therapy or failure to treat with conventional therapy;
- the consequences of exposure of some subjects to possible harm while depriving others of possible benefits;
- acceptability of introducing new treatments into routine use without prior testing;
- extent to which trial is explained to subjects and procedures for obtaining informed consent;
- care for welfare and safety of subjects while preserving "blind" assessment.

Cross-over trials

Each group receives both treatment and placebo but in random sequence. Sometimes this occurs in a RCT by default owing to subject choice or for medical reasons. In this case, to avoid bias, the outcome must be analysed with all subjects in their originally assigned group (classification by "intention to treat").

See the companion text for practical exercises on case-control and cohort studies, clinical trials, and comparing different study designs.

CHAPTER 5

THE SUBJECT OF RESEARCH: SELECTING A POPULATION

5.1 Defining a population to study

The population from which study participants are to be drawn, the target population, must be carefully selected in relation to the purposes of the study. The sample population, consisting of those who are chosen from the target population for study, must be selected in terms of how representative it is and how feasible it is to access.

Research question truth in the universe		**Study plan** truth in the study
Step 1 **Target population**	**Step 2** **Accessible population**	**Step 3** **Sample population**
Specific clinical and demographic characteristics	Specific temporal and geographic characteristics	Defined approach to sampling

Criteria for Selection

Suited to the research question	Representative of target population Easy to study	Representative of accessible population Easy to do

5.2 Selecting a sample

Carrying out studies in entire populations is nearly impossible. Because of this, as a practical alternative, studies are usually carried out in a sample of available individuals, the accessible population, chosen to represent the total population from which they are drawn, the "target population". Estimates of the frequency of relevant variables present in the target population can be derived from the sample selected if it is appropriately drawn.

Studying an entire population (the target population) is costly and presents logistical problems. For this reason, a sample of the population (the sample population) is chosen on whom the measurements are performed. Frequently, it is not possible to select the sample from the whole population so a sample is selected from that part of the population that can be accessed (the accessible population). If the accessible population does not differ in any appreciable way from the target population (the distribution of key variables is shown to be the same), the sample drawn from the accessible population likely reflects the characteristics of the target population. In this case, inferences can be drawn, from measurements on the sample, to the target population. Should the accessible or the sample population differ appreciably from the target population, this is not possible.

5.2.1 Definitions and procedures

In sampling, the whole collection of units (the "universe") from which a sample may be drawn is the population. It is not necessarily a population of persons; the units may be institutions, records or events. The sampling process is intended to yield a population for study that is representative of the whole population. The units to be used when selecting a sample must be specified as this determines the approach to be taken to sampling as well as to analysis.

The target population is the collection of individuals (items, measurements) about whom we want to draw conclusions (make inferences). The population about which inferences are made may be the single population from which a sample is drawn (accessible population, for example, a sawmill) or sometimes all similar "target" populations (for example, all sawmills) about which inferences are required.

The sample (study) population is the group of individuals (units) chosen for study from the accessible population. The characteristics of the sample population are intended to be typical (representative) of the target population to which the findings are to be extended.

Where possible, comparisons between characteristics of the target and sample populations should be displayed to identify any possible differences that might bias the results. A full and complete description and comparison of participants and non-participants within the study population should also be made.

Points to consider in selecting populations

Populations are selected by:

- appropriateness: whether or not the study objectives can be completed using this population;
- practicability: whether or not it is possible to do the study in this group for reasons such as accessibility, willingness of the members of the population to participate in the study and the financial costs of the study.

Example:

> In proposing to study the effects on lung function of exposure to wood dust, one might choose to study people who work in a sawmill. A sawmill may be chosen because there are sawmills in the neighbourhood, your brother-in-law is the manager, it represents a group of people easily accessible to study.

The study population should be defined in terms of place, time and other relevant characteristics.

General aspects of selecting a sample

The sample needs to have the following characteristics:

- it is representative of the target population;
- it is sufficiently large to minimise the effects of random variations in the data;
- it is adequate to represent all groups of potential interest.

Example:

> The type of sawmill chosen, in terms of the material handled, must reflect the question posed. If we are looking for asthma caused by wood dust exposure, we must select a sawmill handling the type of wood we suspect causes asthma. From several hundred sawmills in an area, we would choose one with enough workers to reflect the variation in level of dust exposure that is needed to answer the question.

Key decisions in the sampling procedure include:

1. How do I identify the persons to study (what is the sampling frame)? e.g., population lists of censuses, registries, patient lists, employment lists;
2. How do I ensure that the group studied is representative of the target population and not a biased sample that would lead to incorrect conclusions?

3. What is the size of the sample to be studied? Is the number large enough to ensure that the study can answer the question posed (i.e. has adequate statistical power to reject or accept the null hypothesis chosen for the study). Relevant considerations include:
   - Accuracy of measurements (intrinsic variability);
   - Degree of precision required;
   - Magnitude of differences between groups that are regarded as of practical importance;
   - Available resources.

From where do I select a sample for study (what is the sampling frame)?

The usual means of selecting study populations include:

- Population-based samples. These may be drawn from population registers, from census data bases or from direct contact methods, such as telephone sampling;
- Institution-based samples. These may be drawn from the work place (occupational lung disease studies), from professional associations (studies of lung cancer and tobacco smoking), from schools (the classic sampling frame for tuberculin surveys) or from lists of health services users (studies of asthma, tuberculosis or tobacco use);
- Other approaches. These are often convenience samples or other haphazard methods that may be subject to unknown selective biases.

5.2.2 How do I select the sample?

Commonly used methods are:

Random sampling, when each sample unit has the same probability of being selected (for example, using a table of random numbers).

Systematic sampling, in which subjects are selected at regular intervals from a list (for example, choosing every fifth person on a roster of factory employees).

Cluster sampling, which consists of a simple random sample of clusters of individuals (for example, households)

Stratified sampling, when the population is divided into subgroups or strata (for example, by age group, gender or ethnic origin) and separate random samples of varying size are drawn from each stratum. In this way the representation of smaller groups can be increased, (e.g. by sampling 1 in 2), while that of larger groups is decreased, (e.g. by sampling 1 in 5).

Multi-stage sampling which is a combination of two or more of the above methods.

5.2.3 How do I ensure that wrong conclusions are not drawn (Validity of the study)?

Selection should be carried out in such a way as to avoid incorrect (spurious) conclusions at the end of the study. There are three different levels of accuracy that can be affected by the sample selected:

- Precision:
  The size of the study population affects the precision of the statistical inferences. This is discussed in detail in subsequent sections of this text.
- Internal validity:
  This refers to the accuracy or lack of distortion (bias or systematic error) of the estimated measure compared to its true value. It can be influenced by factors other than the sampling variation, such as the characteristics of the population selected. The selection process itself can result in systematic errors.
- External validity:
  This refers to the generalisation of the results to populations other than the sample studied. Any such extrapolation should always be made with caution. The results of studies that are internally valid cannot necessarily be generalised to other populations.

Validity problems arising from bias in the selection of populations (called **selection bias**) may produce major distortion of results greater than those arising from statistical inference. More detailed discussion of this issue is included in the subsequent text.

See the companion text for a practical exercise on sampling a population.

5.3 Deciding how big the study population should be

Before any research project is started, it is essential to make sure that the size of the study sample will be large enough so that the researchers can be comfortable that conclusions drawn from the study results are likely to be true. On the other hand, it is also important to ensure that resources are not wasted (or too many people inconvenienced) by having a larger study sample than necessary.

Example:

To study the relationship between passive smoking and lung cancer in non-smoking women, a group of non-smoking nurses is followed for 30 years and rates for lung cancer calculated for those exposed to passive smoke at home or work and those not exposed.

Before the study is started, it is important to ensure the sample of nurses will be large enough to be confident that if a difference in lung cancer is seen after 30 years, the difference is "true"' and not simply the result of random variation. On the other hand, since the study will be costly and lengthy, it would not be efficient to enrol many more nurses than necessary to answer the study question.

5.3.1 Why is the size of the study important?

Most research seeks to compare disease rates between groups, or to evaluate associations between determinants and disease. To do this, researchers study or observe a sample of the target population, and draw conclusions about "truth in the population" or "truth for other similar populations" based on results from the sample population studied.

However, results from the sample may not always reflect the "truth" about relationships between disease and determinants in the target population because of both random and non-random (or systematic) variation in the way subjects are selected or measurements made.

It should be obvious that the larger the study, the smaller the chance of making an erroneous conclusion because of random variation or error in the results. Similarly, if the true difference in disease rates between groups is very large, it should be possible to detect the difference with a relatively small study. However, if the true difference in disease rates is small (but real), it will take a larger study to be certain that the results reflect the truth. Finally, if there is truly a difference in disease rates between two groups, but our measurement of disease is subject to a large amount of error, a larger study will be needed to ensure that differences found in the study are not incorrectly attributed to error.

See the companion text for a practical exercise on random variation and sample size.

5.3.2 Information needed to determine the size of the study

The following must be taken into account in order to determine the size of the study:

Null hypothesis

Sample size calculations are specific to the hypothesis being tested, therefore, the null hypotheses must be clearly stated. Often, the sample size calculation is carried out only for

the major hypothesis of the study. Alternately, it can be done for the hypothesis that will be tested using the smallest subgroup in the study.

Example:

> Null hypothesis (H_0):
> There is no association between passive smoking and lung cancer (or the lung cancer rate among non-smokers exposed to passive smoke and those not exposed is equal).

The certainty about not drawing an erroneous conclusion from the results

The notion of "certainty" comprises two different concepts, illustrated in the table and examples below.

Possible conclusions based on results from a study comparing lung cancer rates in these two groups are displayed in this table:

		Truth about the population	
		Passive smoking IS related to lung cancer	Passive smoking is NOT related to lung cancer
Conclusion, based on results from a study of a sample of the population	Reject the null hypothesis (i.e. rates in the study appear to be different)	**OK**	**Type I Error Probability =** α
	Accept the null hypothesis (i.e. rates in the study appear similar)	**Type II error Probability =** β	**OK**

Two types of erroneous conclusions are illustrated:

1. If lung cancer rates appear to be different in the study but in truth there really is no biological association, you will "reject the null hypothesis" incorrectly (Type I error).

   Alpha — the probability of making a type I error (i.e. concluding that differences or associations are real when they are actually not real);

   Confidence level — how certain you want to be that you don't make a type I error ($1 - \alpha$).

2. If lung cancer rates appear similar in the study, but in truth, passive smoking really does increase lung cancer, you will "accept the null hypothesis" incorrectly (Type II error).

Beta	the probability of making a type II error (i.e. concluding that there is no difference or association when there really is one);
Power	how certain you want to be that you don't make a type II error ($1 - \beta$).

Expected study results (effect size)

In order to estimate how many study subjects are necessary to test the hypothesis with the determined levels of confidence, information is needed on the expected size and variability of the study result! This is called effect size.

Effect size depends on 2 things:

- how large is the expected difference between the groups, and
- how much variability is expected in the measure you plan to use to evaluate the groups.

In other words, it is necessary to estimate the size of the "signal" in relation to the "noise".

5.3.3 How do you do a sample size calculation using these factors?

It should be apparent that sample size, study power, confidence level, and effect size are inextricably linked to each other. To know (or be able to estimate) any one of these parameters, it is essential to know (or estimate) the other 3 factors. For example, to estimate the necessary sample size for a study, the research team must decide, in advance, the confidence level, the acceptable study power, and the expected or desirable effect size.

Sample size may be the factor you wish to calculate. Alternatively, it may be fixed by the number of patients or subjects available for study or it may be relatively fixed by the available money, time, or other resources. In this case, it is power, confidence or effect size that needs to be determined.

Estimating the values for the factors in the equation

How to choose confidence level (or 1 – alpha)

- up to the research team to decide this (or this may be the factor you wish to calculate);
- often, by "default" researchers use alpha=0.05 (to give a 95% confidence interval);
- choice depends on the danger associated with making a type I error.

Example:

> If a new approach to treatment has serious side effects, the researchers would want to be very confident about declaring that the treatment makes a difference (i.e. is better) over existing treatments. In this case, one might choose alpha=0.01 to give a 99% confidence interval.

How to choose power (or 1 – beta)

- up to the research team to decide this (or may be the factor you wish to calculate);
- often, by "default", researchers choose power = 80% (i.e. to be 80% confident of not making a type II error). The usual reason for choosing this value is that researchers often feel that making the mistake of saying there is no difference (e.g. between two treatments or two groups) when there really is one, is "safer" than the reverse mistake;
- choice depends on the danger, and the cost, of making a type II error.

Example:

> If the new treatment is much cheaper and faster, and the researchers only have one chance to study the effectiveness of the drug in their setting, they would want to make sure that they don't make a mistake and declare the drug not effective, if it really is effective. So, they would want to design a study with a higher level of power (say 95%).

Information to estimate effect size can come from:

- other, similar studies;
- a pilot study;
- an informed "best guess" or an estimate of an "important difference".

This is often a very difficult thing for researchers to accept. It is common for a researcher to consult a statistician for help with a sample size calculation only to be surprised when the statistician insists on knowing the expected average difference in disease rates between the groups, or the expected variance of the measure being studied.

The researcher is tempted to reply "if I knew the difference I would not need to do the study!". Unfortunately, the only way to estimate the appropriate sample size is by making an educated guess about the "effect size" or to choose an "effect size" which has a meaning in terms health services action.

Formulae for calculating sample size (or confidence, or power, or effect size)

The formulae used most frequently for sample size/power are those which relate to comparing disease rates (e.g. prevalence of multi-drug resistant tuberculosis) or comparing average values of a measured variable (e.g. FEV_1) between two groups. If your study protocol calls for other types of statistical analyses, you should discuss sample size/power with a statistical consultant before starting the study.

1. Comparing disease rates between 2 groups:
   (e.g. in a cohort, or cross-sectional, or experimental study):
   EPI-INFO has a statistical calculator that will calculate sample size (or power, or effect size), if you provide values for the other factors.

   The formula used is:

$$\text{sample size} = \frac{[(p_0 \text{ X } (1\text{-}p_0)] + [(p_1 \text{ X } (1\text{-}p_1)] \text{ X } (z_{(1\text{-alpha})} + z_{(1\text{-beta})})^2}{(p_1 - p_0)^2}$$

   Where:
   p_0 = expected rate of disease in the non-exposed group (or group 1)
   p_1 = expected rate of disease in the exposed group (or group 2)
   $z_{1\text{-alpha}}$ = z-score (see table) for the alpha you selected
   $z_{1\text{-beta}}$ = z-score (see table) for the beta you selected

   note about z-scores:
   Most statistical software packages ask you simply to input the values for alpha and beta, the software converts this information to numeric values (called z-scores) depending on the type of study being proposed. If you want to do the calculations without a computer, you will need to read off the appropriate z-score from a z-score table found in most statistics textbooks.

2. Comparing exposure (or risk factor) frequency between two groups:
   (e.g. case-control study)
   The formula used is identical to the one above, but in this case p_0 and p_1 refer to the frequency or proportion of the exposure, or risk factor in the non-diseased and diseased groups.

   Comparing mean values for a measured variable between two groups
   (e.g. cross-sectional or experimental study)

$$\text{sample size} \quad \frac{2(SD)^2 \text{ X } (z_{(1\text{-alpha}/2)} + z_{(\text{beta})})^2}{(\text{mean}_1 - \text{mean}_2)}$$

   Where: SD = expected standard deviation of the measure in the population

3. Sample size for analyses that will use "correlation" or regression:
   It is possible to compute an estimate for sample size for hypotheses that will be tested using correlation or regression analysis. A statistician should be consulted for this. However, a useful "rule of thumb" can also be applied to give an estimate that is useful for study planning purposes. The "rule of thumb" is that one needs about 30 subjects for each different factor to be examined using regression modelling. For example, if you plan to study lung function (e.g. FEV_1) and need to take into account age, gender, history of asthma, socio-economic status, smoking history, and a measure of environmental exposure (6 factors) a fair starting assumption would be about 6 X 30 or 180 subjects.

A few words about one-tailed and two-tailed tests in using sample size formulae

In the calculations described above the values for the z scores are found in standard z-score tables. The z-score for alpha assumes that $^1/_2$ of the possible type 1 error is distributed at one extreme of a normal distribution and the other $^1/_2$ is at the other extreme of a normal distribution. This is called a 2-tailed test.

A 2-tailed test is usually the correct one, and unless there are very specific reasons why no new knowledge would be gained from a "negative" result, a 2-tailed test should be used.

Further detail about one-tailed and two-tailed tests (optional)

Most of the time when a research project is conducted the investigator has a hypothesis about what the results might be. For example, in a study of risk factors for asthma, the researchers may hypothesise that urban air pollution is associated with an increase in the rate of asthma attacks. However, if the study indicated that people living in highly polluted urban environments had significantly fewer asthma attacks, that would also be an interesting finding. In other words, regardless of the initial hypothesis, in most research, results are informative no matter which "direction" they follow.

On occasion, the researcher is ONLY interested in results that fall in one direction.

For example, you are planning to evaluate a new treatment for TB that would be shorter in length and require fewer visits. This study has three possible outcomes: (1) the standard treatment is better; (2) there is no difference between the two; (3) the new treatment is better. However, in this case, outcomes (2) and (3) are functionally equivalent since either would lead you to adopt the new treatment. Put another way, we have no need to distinguish between outcome (2) and (3).

In this example, when calculating power or sample size, it would be appropriate to use a one-tailed test, since you are NOT interested in the possibility that the new treatment is better – you are only interested in knowing if it is worse. In this case, you select the z-score for alpha/2. When you do this, it is important to remember that the power for the analysis of the opposite effect (i.e. that the new treatment is better) is zero.

See the companion text for a practical exercise on using sample size and power calculations in developing study protocols.

CHAPTER 6

MEASUREMENT IN EPIDEMIOLOGY

6.1 General principles

6.1.1 Principles of collecting information (data)

The methods for data collection and the criteria for entry of each variable into the research records must be precisely defined at the outset of the study and rigidly applied throughout. Any later changes should be resisted but if made the reason for making the change must be clearly documented. The criteria should refer to any factor of potential significance to the study, such as time, place, age, gender and ethnicity.

The usual result of measurement is to produce numbers or create categories. These numbers or categories are arranged in such a way as to identify patterns from which logical conclusions can be drawn. Their relation to one another tells us (we hope) about the world around us.

6.1.2 Types of data

The data or information to be collected should be specified in the study protocol in order that relevant analyses can be made. There are two main types of data:

Continuous data

These are points located on a continuous scale of values (for example, height or age and some measures of function such as FEV_1). The numbers reflecting continuous data have no "discrete" reality and flow into one another along a scale.

Example:

> Although we say someone is 52 years of age, this is not absolutely accurate. It only "approximates" the age, to the nearest year. Adding decimal points (for example, 52.69 years) can indicate the actual value more precisely. Increasing the number of decimal points increases the "precision" of the estimate. Because it is not possible to be "absolutely" precise, such measurements are usually recorded within "intervals" (that is to say, someone is 52 years of age, with an age lying between 52.00000….. and 52.999999…..).

Discrete (categorical) data

These represent groups of unique occurrences. Information as to whether an individual or object falls into one or other category is usually indicated by an answer of "yes" or "no" to that category. Thus, for gender, one is either male or female; there is no other category. The categories are mutually exclusive and do not overlap. If you are not female, you are certain to be male. There may be more than two discrete but independent possibilities, for example, number of children in the family.

Death may be considered an example of discrete data; one is either alive or dead and it is usually straightforward to determine into which category an individual falls. When it comes to the cause of death, however, measurement is less precise.

Most "diseases" are difficult to categorise using discrete variables. For example, it is assumed for categorisation purposes that one either has or has not got a disease such as asthma or tuberculosis or pneumonia (i.e. they are discrete data). However, the boundary between having and not having any one of them is far from clear and what one person might call pneumonia another might not. Categorisation is, therefore, usually based on arbitrarily assigned criteria in standard definitions.

For this reason, research protocols must set out as precisely as possible the definitions of the states (diseases) or characteristics (agents/factors/determinants) they are proposing to study. Failure to do so will lead to lack of precision of measurement and will confuse the investigators and jeopardise their ability to draw conclusions from the information collected or reported.

6.1.3 Instruments to measure states or characteristics (collect data)

Information is gathered using tools or instruments. In studying lung health, the types of tools may be such things as:

- interviews, questionnaires, diaries, forms and routine records;
- measuring devices such as rulers to measure skin reaction to tuberculin or aeroallergens, spirometers, devices for sampling ambient air pollution;
- x-ray films;
- microbiological techniques;
- techniques for evaluating tissue samples.

Some instruments are more precise than others are. This is usually the case where the categories have no overlap and the decision into which category an individual falls is relatively straightforward.

Example:

> If we wish to collect information on gender, we may look at the individual, we may ask the individual to specify gender or we may record a gender category from records of one kind or another. In most instances, it is fairly easy to distinguish between male and female gender.

In contrast, when categories are not truly discrete and there is significant overlap between categories (e.g. between those who do and those who do not have a disease such as pneumonia) or the instruments used to measure them are subjective (for example, reading chest radiographs), measurement is considerably less precise. This is also an inescapable feature for measured values, where a "judgement" has to be made concerning the interval into which a single measurement may fall (e.g. measuring induration in response to a tuberculin skin test).

A number of standardised and validated methods have been routinely used in studies of lung health. Where possible, these definitions and methods should be followed precisely to ensure comparability with other scientific studies previously reported. Where departures are made, these should be fully described.

6.1.4 Problems with measurement

The information proposed to be collected, the methods by which it will be obtained, and how it will be categorised must be precisely defined. This is to ensure that there is no ambiguity in the data recorded and that other investigators or critical reviewers are able to confirm your results or compare them with those obtained in other already published studies.

Ensuring comparability: standardisation of definitions and procedures

Standardisation is a process whereby methodological differences between studies can be minimised. It is applied, for example, to diagnostic definitions or to instruments and procedures used to collect data.

The definitions chosen for study purposes usually conform to those recommended for comparison purposes. In topic areas where research is extensively carried out, these definitions are often established by international consensus, based on scientific work already published. Where such international recommendations exist, investigators must either follow them or indicate the precise manner of departure from the recommendations and the rationale for the departure.

The same is true for techniques and instruments of measurement. If a standardised, international recommendation exists concerning a procedure or type of instrument for measurement, these should be the preferred methods and need to be precisely described and followed. Any departure from these recommendations must be fully described and justified.

Ensuring precision: minimising error in measurement

Errors in measurement may reflect carelessness on the part of study personnel; inadequacy or incorrect use of measurement tools and instruments or failure to follow standard procedures aimed at minimising such errors. As will be discussed in subsequent sections, error may be random or systematic.

As much as possible, the actual values of data collected should be reported, as this gives an indication of its quality. The presentation of "real values" is surprisingly frequently not done.

Example:

> A classic example of error in measurement is "terminal digit preference". This is illustrated in tuberculin skin testing where, in the absence of attention to careful technique, research (or health care) personnel tend to preferentially read certain results as multiples of five. Where care is taken to improve precision in measurement, such errors can be avoided. Indications of terminal digit preference are indicators of the quality of data collection.

Precision in measurement is often a personal characteristic of individual research personnel. To ensure the greatest precision of measurement, the measurements taken by each of the research personnel should be compared with the others and with a "gold standard" (inter observer error). If the measurements recorded by an individual research

worker vary greatly from those of the others, either the worker must be trained until the variation is reduced to a minimum or the worker must be excluded from making the measurements. In addition, measurements of the same values should be taken more than once by each of the individuals meant to carry them out in a research study, to ensure that the measurements taken on different occasions do not vary greatly from one another (intra observer error).

Completeness and accuracy of recording

It surprises many inexperienced investigators that the process of recording and compiling the measurements made may contain many errors, sufficient to destroy a study. No matter how careful personnel are in carrying out measurements nor how precise the instruments in providing information, if the information is not carefully and completely recorded andaccurately compiled, the information is no longer of value for scientific purposes. No amount of statistical manipulation will overcome problems created by sloppy handling of information.

Even after precise and careful measurement and even with careful attention in the recording and compiling of data, the amount of error introduced in the stage of recording and compiling may be surprisingly large. In many good studies, this error may be as high as several per cent of the entire information in the study. Great care must therefore be taken to minimise this error to the greatest extent possible and to estimate the amount of it in any given study. This will be discussed further in sections dealing with data management.

6.2 Counting disease and measuring health

Health may be seen as a continuum and, as such, can be measured in terms of functional capacity (the degree to which one is able to carry out given tasks or activities). Disease impacts on health in that it inhibits functional capacity and can jeopardise the life of the individual. Thus, the state of health can vary from perfect functional capacity to extreme dysfunction and death. The various spheres of measurement used to categorise state of health include death, disease incidence and prevalence, functional capacity, pre-morbid states, risk group.

6.2.1 Sources of information

Information on health and disease may be obtained from a variety of sources. Information frequently used for epidemiological purposes is that which is routinely recorded in the course of care of patients or for disease surveillance or monitoring in a population. Alternatively, but much more costly, information may be collected specifically for study purposes.

To be useful for scientific purposes, information must be relevant to the subject under study, reliable, complete and accessible.

Routinely collected information

Information routinely recorded in the care of patients may be used for scientific purposes. This type of information, while usually relatively accessible and relevant to the subject under study, is frequently neither complete nor reliable. It is a sobering experience, even for an epidemiologist, to review a series of records of cases under his/her care in routine practice to try to extract information for purposes of a study.

Example:

> It is instructive to review routine records for details of the tobacco smoking habits of patients. It is striking how often this information is incomplete, even though we know this is one of the most important factors associated with lung disease.

Because of these problems, the most useful routinely collected information is that which is collected on a standardised form. Such forms are frequently internationally recommended (for example, the information forms recommended in the series of "Guides" published by the IUATLD on Tuberculosis, on Acute Respiratory Infections in Children, on Asthma and on Tobacco Control and Prevention). When such forms are regularly used, information is more likely to be complete and comparable from one patient to another.

Such forms are also used for official statistics such as registration of vital events (deaths, births, and marriages) and of diseases subject to mandatory notification (such as tuberculosis).

Information collected specifically for study purposes

Frequently, routine information relevant to a particular subject is not available or is not of sufficient quality to be useful. In this case, the information must be collected specifically for purposes of the study. This is clearly more expensive but is frequently the only way to obtain reliable and complete information.

As in the case of routinely collected information, that which is collected using standardised, field-tested forms is the most reliable for study purposes.

6.2.2 Studies of mortality

Death is the state easiest to define along the continuum of health; defining its cause is wholly another matter.

Recording death

Information on deaths is usually obtained from routine notification of deaths to the Vital Statistics Register in a given locality. This information is collected using a standardised form and classification of causes of death follows an internationally recommended format.

A physician caring for the patient usually completes the form, although in some instances, a physician or coroner who has no prior knowledge of the patient completes it.

Measuring death

All rates referring to deaths are "incidence" rates as death is a discrete and time-limited event. Death rates are used routinely to measure lung diseases (such as lung cancer) with a high case-fatality ratio. Where interventions are effective (such as in tuberculosis or pneumonia) death rates are more of an indication of the quality and accessibility of health services rather than the burden of disease in the community.

Death certificate

Cause of death		**Approximate interval between onset and death**
I Disease or condition directly leading to death*	(a)	.....
	due to (or as a consequence of)	
	(b)	.....
Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last	due to (or as a consequence of)	
	(c)	.....
	due to (or as a consequence of)	
	(d)	.....
II Other significant conditions contributing to the death, but not related to the disease or condition causing it	.....	.....
	.....	.....
* This does not mean the mode of dying, e.g. heart failure, respiratory failure. It means the disease, injury, or complication that caused death.		

In measuring mortality, various rates are used:

- Crude death rate indicates the probability of death in a defined population in a specified period;
- Cause- (or age-) specific death rate is the probability of death from a given disease or condition (or in a specific age group) in a defined population in a specified period;
- Case fatality rate is the probability of death from that disease or condition in individuals with the given disease or condition within a specified period of time.

The following are the calculations used to derive the rates in a defined population:

Expression	Usual units of expression
Crude mortality rate: $\frac{\text{Number of deaths during the period}}{\text{Persons at risk of dying during that period}}$	per 100,000 per year
Cause-/age- specific mortality rate: $\frac{\text{Number of deaths from a specific cause/in a specific age group during the period}}{\text{Persons at risk of dying during that period}}$	per 100,000 per year
Case fatality rate: $\frac{\text{Number of deaths from a specific cause during the period}}{\text{Number of cases with the disease during that period}}$	per cent

People with a disease may die of causes other than the disease itself.

Example:

> The most obvious example of this situation is an aged individual, for example, 91 years old. The probability of remaining alive for the coming year is usually much less than that of an individual who was 21 at the beginning, irrespective of a particular disease or condition the individual may have. In comparing one population with another, therefore, it is necessary to account for those other factors (age, gender, socio-economic level) which may increase the probability of dying if a person has the particular disease or condition.

6.2.3 Studies of Morbidity

Morbidity is any departure, subjective or objective, from a state of physiological or psychological well being. Usually this term refers to a disease or to a functional state that is precisely defined. Morbidity can be measured in three ways:

- persons who are ill (cases);
- periods or spells of illness that the persons experience (events);
- impact (duration, extent of functional impairment, service utilisation) of illnesses.

Features to be considered when counting diseases and measuring health

Frequency

The frequency of occurrence of a disease is an important factor in selection of study design. The study of rare diseases (lung cancer, tuberculosis) is less efficient using a cohort design. The case control design may be more practical.

Severity

In comparing determinants of disease or function between groups, the severity of disease or of dysfunction must be considered.

Examples:

- semi-quantitative evaluation of degree of positivity of sputum smears which is associated with the degree of infectiousness of a case of tuberculosis;
- extent of involvement of the lungs which is associated with survival and response to treatment in tuberculosis;
- level of function and frequency of symptoms (severity grade) which is associated with health care utilisation in asthma;
- level of physiological function which is associated with disability and survival in chronic airflow obstruction.

Duration

The duration of disease varies widely in lung diseases. Acute respiratory infections are usually of very short duration while asthma is occasionally life long. The duration of disease has an important influence in determining the selection of study design. Diseases of short duration cannot be studied efficiently using cross-sectional study designs.

Latent period

Many chronic diseases caused by environmental exposures (chronic airflow obstruction, cancer) and infectious diseases with a carrier state (tuberculosis, community acquired bacterial pneumonia, fungal diseases and cytomegalovirus disease) have a latent period. Where a latent period is a feature of disease, it must be taken into account in the design of the study.

Choosing an instrument or method

Instruments for recording routinely collected information on cases, events or impact

Of the various lung diseases, recording and reporting of cases and morbid events in most countries is carried out routinely only for tuberculosis. In this case, there is an international, standardised system for recording and reporting with agreed definitions and procedures. Training materials are available to ensure that the quality of recording and reporting is acceptably standardised. In addition, certain indicators have been identified which allow a general assessment of the quality of the information that is reported. The recommended definitions and procedures form the framework within which tuberculosis research is carried out.

Routine information on impact in the form of hospitalisation for asthma and for acute respiratory infections is also reported. This information is much less well standardised and is more difficult to use for research purposes.

Systems for routine recording of cases, events or impact of disease have certain technical requirements if they are to be used in research. These are:

- the records must be simple, containing the minimum of information required to ensure accuracy;
- they should be part of routine practice and useful to the practitioners to encourage completeness of reporting;
- there must be agreed operational definitions with systematic training of those using the forms to ensure consistency of response;
- there must be a functioning system of communication to ensure that the information is regularly submitted and the results or relevant consequences fed back to those completing the forms;
- a routine system of monitoring of the completeness and accuracy of information (including agreed indicators and regular supervision) is necessary for sustainability.

Instruments for recording information collected specifically for study purposes

Collection of data specifically for research purposes is frequently necessary and a series of standardised definitions and procedures is available for doing this work.

Features of methods or instruments to consider when making the choice include:

- the existence of standardised, recommended methods for research;
- the accuracy (test characteristics) of the instrument or method to be used;
- the durability of the instrument to be used;
- its acceptability to the participants;
- potential hazards associated with its use;
- maintenance of confidentiality.

6.3 Use and design of a questionnaire

Questionnaires are frequently used as instruments for the recording of information (data collection) in epidemiological investigations. Good design is essential to gather accurate and reliable information.

The aim is to ensure that as far as possible information is obtained in a systematic manner in accordance with pre-determined and standardised criteria and methods. This allows valid comparisons to be made between the attributes of different groups of study subjects.

Questionnaires may be administered by face-to-face interview or self-completed by the study subjects. The mode of completion will affect the format but the principles remain the same.

It is always important that questions are precise and unambiguous and that clear instructions are given to the interviewer or the subject, as appropriate, on how to interpret questions, how to record answers, and how to proceed through the questionnaire.

These notes are intended to give guidance on good practice in questionnaire design.

6.3.1 Principles of questionnaire design

Aims of the Study It is essential that the aims of the study are clear and precise.

Question Selection All questions necessary to fulfil the aims of the study should be included. Equally questions that are not directly relevant to the aims must be rigorously excluded.

Question Wording The wording of questions must be intelligible to all subjects and unambiguous.

Question Sequence The order in which questions are asked needs to be carefully planned. It should be logical, interesting and economical

Response Options The types of response being sought must be explicitly set out (i.e. discrete options or open-ended answers).

Questionnaire Structure The structure and layout of the questionnaire must be designed to facilitate its completion and analysis of the data.

6.3.2 Content of the questionnaire

Briefing

This should provide a short statement indicating:

- the auspices and purpose of the study,
- how subjects (cases and controls) have been selected,

- the procedure being followed (interview or self-completion),
- the rights of subjects to decline to participate or to answer specific questions,
- arrangements for safeguarding confidentiality and,
- the subject's signed consent to participate, when appropriate.

Identification: subject's name and address, and subject identification number

The information gathered in this section should be recorded on a cover sheet that is then detached and kept in a file separate from the remainder of the data collected. The subject identification number alone is used to label all other data. The connection between personal identifiers and subject identification numbers must be securely filed such that only the investigator has the ability to access both. This assures the confidentiality of data on identifiable individuals. It is unethical to hold the information connecting the personal identifiers and the subject identification number in any file (hard copy or electronic) that has any risk of being accessed by anyone without authorisation.

Personal characteristics

The data recorded will depend on the nature and purposes of the study, but usually includes:

- title of survey and subject identification number;
- date, time and place of interview and name of interviewer;
- age, gender, marital status, occupation, social data relevant to the purpose of the study (e.g. education, income, housing, family size).

N.B. Data on personal characteristics such as these are often useful to check for bias in subject selection or participation by comparing the frequency of specific characteristics in the study population with those in the target population.

Question Selection

All questions must be strictly relevant to the aims of the study. There is no place for superfluous questions; for example those included on the basis of "while we have the chance it would be nice to know"! Similarly, for reasons of economy and to sustain the interest and co-operation of subjects, the number of questions should be kept to a minimum consistent with the need to ensure that all the information required to answer the study question is obtained.

On occasions, however, it may be economical to combine more than one topic into the same survey. In such cases the objectives of each component study must be equally clear, the level of participation must not be compromised by overburdening participants, and the same degree of rigor must be applied to the selection and composition of questions for inclusion in relation to each topic.

Question Wording

The wording of questions is crucial to obtaining good quality information that can be relied on to serve the study purpose. They should have the following properties:

- simple - questions should be short and uncomplicated, asking only one piece of information per question;
- intelligible - as far as possible use words that are within the normal vocabulary and understanding of the subject (in some cases a short explanation of technical terms or an alternative phraseology may be helpful and/or guidance to interviewers on supplementary probing questions and interpretation of answers may be advisable);
- unambiguous - questions should not be capable of more than one meaning (e.g. "are you on drugs"? - requires definition) and must be precise (e.g. "do you cough often"? - requires specification of frequency);
- unbiased - avoid leading questions that imply expectation of a particular answer (e.g. "do you think that smoking is bad for you"?). Avoid the use of emotive words (e.g. "do you think taking drugs is immoral"?) Do not encourage acquiescence to fashion or what is thought to be socially acceptable (e.g. "do you think all people with TB should be isolated"?);
- interest - questions should as far as possible seem relevant and interesting to subjects in order to sustain their concentration and willingness to participate.

Where subjects whose first language is not English are involved, the questionnaire should be translated by one linguist and then translated back by a different linguist. The result should be compared with the original to ensure that the meaning has not been distorted in the translation process.

Question Sequence

The order in which questions are asked may crucially affect the answers given and needs careful thought in order to avoid tedium, embarrassment and biased answers. Thus, it is important that:

- Subjects feel at ease with the questions they are being asked. It is usual, therefore, to start with questions that are easy to answer, such as identifying and demographic data which will build the subject's confidence. Questions about personal habits and lifestyle, for example, which may be more threatening or embarrassing, should be reserved till later in the questionnaire.
- The sequence of questions should flow easily and logically and be of interest to the subject as well as to the investigator. Any very complex questions are best deferred until late in the questionnaire.

- The answers to early questions should not prejudice the answers to later questions. For example, questions about the harmful effects of smoking might influence subsequent answers to questions about the subject's own smoking habits.
- "Check questions", i.e. those designed to check the consistency and thus the validity of answers (e.g. "age" and "date of birth") should be well separated.

Response Options

The response format must give subjects scope to provide accurate and complete answers. There should be on offer, therefore, an appropriate menu of response options. There are three general types of response format:

- "Yes", "No" (or "Don't know"). The main disadvantage of the dichotomous answer is that shades of meaning cannot be distinguished.
- Multiple choice in which several different possible answers are listed from which the subject may choose one or more options, including "Don't know". It is often difficult to be sure that the list caters for all possible responses and therefore an option of "Other….specify" is usually offered. However, while this may be useful as an "escape" option for subjects, any answers have the disadvantages of the "open-ended" format and should, if possible, be avoided.
- The "Open-ended" format is commonly used when exploring uncharted territory such as personal beliefs and attitudes, so-called ethnographic studies. They have the advantage of being free from interrogator bias and allowing subjects to express their views in their own words, often as a prelude to more structured study. The main disadvantage is the difficulty of analysis and reaching general conclusions.

Questionnaire Structure

A good questionnaire should be easy to complete and easy to analyse.

Ease of completion is facilitated by:

- Clear instructions to the interviewer or, in the case of self-completed questionnaires, the respondent. These will include general instructions at the beginning and interval instructions relating to particular questions where necessary;
- Numbering of questions for reference purposes;
- Indication of where questions may be skipped when, for instance, a negative answer renders some ensuing questions redundant.

Ease of analysis is facilitated by:

- Discrete questions and answers which are always easier to analyse and interpret than qualitative questions and answers;
- Use of "boxes" for answers which can be numbered and pre-coded.

- Aligning boxes in margin to facilitate coding and computer entry;
- Ensuring that there are boxes for all possible optional responses to each question, including "Don't know", "Not applicable" and "No answer".

Finally:

- Always use good quality paper and print;
- Use of coloured paper may be helpful to distinguish different questionnaires being used for different purposes in the same study;
- Short questionnaires will seem less daunting to subjects and are more likely to achieve a high response rate.

6.4 Measuring exposure

This section uses exposures to occupational and environmental hazards as an example of measuring determinants or risk factors. Most of the principles outlined here apply to measuring other risk factors or determinants.

6.4.1 Recording or measuring environmental exposure

Just as information about disease can be collected from routine sources or can be information collected specifically for the study purposes, information about environmental and occupational exposures and about other possible determinants (risk factors) for lung diseases (such as socio-economic status, nutrition, co-existing disease) can also come from various sources.

The same provisos apply to this information as discussed above for disease information: using standardised forms to record information, following standardised procedures, and recording as completely as possible the methods used to obtain the data.

Features of exposure (dose) to be considered:

intensity — This refers to the level or magnitude of exposure (e.g. number of cigarettes per day, concentration of silica in the breathing zone of the worker).

duration — This is the length of time over which exposure has occurred (e.g. years employed in a given job; years resident in a region; duration of time spent in contact with an infected person; number of years as a smoker). Depending on the health outcome being studied, the relevant duration to consider may be as long as years (e.g. cancer, chronic obstructive lung disease) or hours or days (e.g. exacerbation of symptoms among patients with asthma).

profile Sometimes it is important to consider whether exposure is roughly constant throughout the exposure period, or if the exposure profile is more a series of peaks separated by periods of no exposure (e.g. repeated accidental spills or leaks from an industrial site).

time period We need to assess the exposure (intensity, duration, and/or profile) at the biologically relevant time period. For example, in a study of environmental or occupational risk factors for lung cancer, it is not helpful to have information about where a person currently works or whether the person currently lives with a smoker. We need to know about exposures at home or at work 10 – 30 years ago.

"setting" vs. agent

In many cases, the relevant specific etiologic agent is not known (e.g. the urban environment, working in an aluminum smelter); in others, the etiologic "agent" is a mixture (e.g. cigarette smoke). We need to keep in mind that exposure can refer to a "setting" or mixture, and need not refer only to a specific agent.

6.4.2 Instruments for collecting exposure information

- interviews, questionnaires, diaries, records;
- measurements of the macro-environment (e.g. level of asbestos in the air in the town, workplace, etc) – called "area" measurements;
- measurements of the personal environment (e.g. measurements taken using an air sampling device worn by the study subject) – called "personal measurements";
- individual dose measurements (e.g. exhaled breath CO, carboxyhemoglobin as measures of cigarette smoking);
- measurements of concentrations in tissue (e.g. quantification of asbestos fibres in lung tissue specimens);
- markers of the direct effect of exposure (e.g. DNA adducts induced by carcinogens in cigarette smoke).

In theory, the further down this list, the more direct or valid the measure of exposure (compared to the "indirect" measures at the top of the list). However, in practice, the limitations of the instruments further down the list often outweigh their advantages.

Advantages and disadvantages of different instruments

Measure	Advantages	Disadvantages
Questionnaires, interviews, diaries, records	• simple • relatively inexpensive • can obtain information from large numbers of subjects • can design the instrument to capture an exposure setting or mixture (e.g. job, area of residence) • can obtain past exposure information and data on exposure duration	• depends on personal reporting (may be biased or inaccurate) • more likely to be qualitative, not quantitative
"area" measurements	• less costly than personal measurements • may be available from routine sources (e.g. air pollution monitoring stations) • quantitative	• may not be a good reflection of the exposure of a given person • measures intensity of exposure at a specific point in time may not be the relevant time (point) • does not capture exposure duration
"personal" measurements	• measures exposure intensity for the person • may be able to evaluate exposure profile • quantitative	• costly, time consuming • measures intensity of exposure at a specific point in time (may not be the relevant time point) • methods may not be sensitive enough (limits of detection) • does not capture exposure duration
Individual dose measures/tissue concentrations	• usually reflects inhaled dose • direct quantitative measure of personal dose	• costs and feasibility vary • may be invasive • need to consider clearance/ washout
Bio-markers of effect	• closest measure of the biologically relevant "effective" exposure	• invasive • often costly • not available for most diseases (e.g. asthma; COPD)

6.4.3 Choosing an exposure assessment method

Choosing an exposure assessment method depends on the available resources and expertise and on the characteristics of exposure that are relevant for the study.

Example:

> Measuring personal exposure of workers in a mine today is not relevant for a study of environmental risk factors for lung cancer, because of the long latent period.

Health professionals often feel that their expertise to measure or record exposure is too limited and that they must rely on other experts for this component of a study, or not include it at all. Although ideally one would consult with a content expert when designing that component of a study that addresses environmental exposures or nutritional aspects, in practice such experts may not be available.

The power of a simple questionnaire (e.g. the American Thoracic Society standardised smoking habit questions or a simple employment history) should not be overlooked. Results from a questionnaire can be used to derive quantitative or semi-quantitative results.

Some semi-quantitative exposure categories:

- intensity and duration of smoking;
- employment in a low/medium/high risk job;
- living above/at/below the poverty line for the region.

Example:

> Self-reported exposure information is often used. In a population-based study of asthma, carried out in 4 Canadian cities using the European Community Respiratory Health Survey protocol, the prevalence of current asthma was 9.6% among over 12,000 adults who answered a mailed questionnaire. Estimating the role of occupational exposure to dust and fumes was done by a simple question: "in your current job, are you exposed to dust, gases, or fumes?" This "exposure" was associated with an almost 2 fold increased risk of asthma (Odds Ratio =1.92, 95% confidence interval: 1.6, 2.2).

See the companion text for practical exercises on defining variables and measuring disease.

6.5 Reporting morbidity or mortality : Calculation of Rates, Indices of Risk

6.5.1 Calculation of rates

In order to make comparisons and draw conclusions, measurements must be transformed into numerical values and then calculated as rates.

Epidemiological measurements refer to specified populations and often to defined periods of time. For example, a rate may be a certain number of health-related states/ events (cases or deaths) per unit of population (usually a multiple of 10 such as 100,000) per unit of time(usually per month or year). These figures allow comparisons to be made between populations (for example, one population has a death rate of 12 and another 55 per 100,000 per year). The expression of rates implies the presence of a numerator and a denominator to obtain a probability or proportion.

Errors and bias may arise in measuring both numerators and denominators.

Numerator errors in routine statistics may arise from variations in:

- use of diagnostic tests;
- case ascertainment procedures;
- recording systems adopted.

Denominator errors may arise from:

- population migration;
- variation in population structure and;
- changes in administrative boundaries.

Prevalence

This corresponds to the number of events of a specific disease or condition in a given population at a designated point in time (point prevalence). It may refer to the number of events over a specified period of time in which case it is designated a period prevalence. When the number of events is divided by the population at risk, it is termed the prevalence ratio (or rate).

Prevalence is most frequently used to measure relatively common chronic diseases with a prolonged clinical course, such as asthma and chronic airflow obstruction.

Incidence

This refers to the number of new events of a specific disease or condition commencing during a given period in a specified population. When this is expressed per unit of time and of population at risk, it is referred to as the incidence rate. The mortality rate is a

special form of incidence rate in that it indicates the rate at which new events (deaths) occur in a population at risk over a unit of time. Incidence rates may be specific to causes as well as to groups in the population (defined by age, gender, race).

Incidence rates are used to measure the frequency of diseases that are either relatively uncommon (tuberculosis) or which have a short clinical course (acute respiratory infections). Incidence rates are frequently estimated by using notification rates. This is appropriate only when case ascertainment is relatively high and an efficient information system is in place.

Service utilisation rates

Service utilisation rates (such as hospital admission rates, length of hospital stay and unplanned visits to health services) are frequently used to estimate the impact of a disease (for example, unplanned visits to the emergency services for asthma). These rates reflect the economic burden of disease and may also indicate the general burden to health in the community. To reflect the latter adequately, services must be both accessible and equitable and reporting needs to be complete and accurate.

Relation of various measures to each other

Each of the measures of disease (and death) is related to each other. Thus:

Prevalence = incidence x average duration of a case;

Mortality = case fatality ratio x average duration of a fatal case x incidence.

Example:

> Where a disease has an average duration longer than one year, the prevalence of the disease will be greater than the annual incidence; where the duration is short (for example, a few days or several weeks), the prevalence of the disease will be less than the annual incidence. When a disease is of long duration (for example, asthma), the prevalence of the disease is more informative; where the disease is of short duration (for example, pneumonia in children), incidence (or mortality) is more informative. In the case of a contagious disease (such as tuberculosis), the prevalence is particularly important, since this reflects the duration of disease, which is a determinant of transmission of infection to susceptible members of the community.

Measures of disease may be affected by medical interventions. Treatment, if it is given properly, may significantly reduce the duration (and therefore the prevalence) of a disease. Effective treatment of any sort rapidly reduces the case-fatality ratio and therefore the mortality rate.

Example:

> Tuberculosis treatment can rapidly decrease the period in which a patient is infectious. However, if it is given improperly, it can actually increase the duration of cases and increase the prevalence (and therefore the transmission of infection) by saving the life of the infectious case without cure.

6.5.2 Indices of risk

Risk is the probability of the transition from one state to another. It may reflect the probability of developing a disease, or of being cured of the disease, related to the presence of a defined determinant or therapy. This probability is the "currency" used to measure disease and its dynamics in a population. Much epidemiological research is directed to the estimation of such probabilities. Results are usually expressed as a rate or ratio per unit of population.

Defining risk

The results of epidemiological studies are often expressed in terms of probabilities or risks. Some commonly used expressions of risk are defined as follows:

Absolute risk

> The observed or calculated probability of an event in a population under study. Indicated by the crude incidence in an exposed population.

Relative risk (RR)

> The ratio of the rate of disease or death among the exposed to the rate among the unexposed (synonym = risk ratio, or rate ratio). Alternatively, the ratio of the cumulative incidence rate in the exposed to the cumulative incidence rate in the unexposed (the cumulative incidence ratio).

Risk difference

> The difference of the rate of disease or death among the exposed to the rate among the unexposed.

Odds ratio (OR)

> The ratio between odds in favour of exposure among cases to odds in favour of exposure in non-cases. The odds ratio is a good approximation of the risk ratio when the incidence of the disease is low.

Attributable risk (AR)

The rate of a disease or other outcome in exposed individuals that can be attributed to the exposure. The measure is derived by the difference between incidence in exposed and incidence in non-exposed.

Population attributable risk (PAR)

The incidence of a disease in a population that is associated with (attributable to) exposure to a risk factor. This reflects the proportion of all the cases in the population that is due to the exposure.

Calculating risk

		Case: Yes	Case: No	Total
Exposed	Yes	a	b	a+b
	No	c	d	c+d
	Total	a+c	b+d	N

Absolute risk	=	a / a+b
Relative risk	=	(a / a+b) / (c / c+d)
Risk difference	=	(a / a+b) - (c / c+d)
Odds ratio	=	(a/b) / (c/d) or ad/bc
Attributable risk	=	(a / a+b) – (c / c+d)
Population attributable risk	=	a+b(AR) / N

CHAPTER 7

CONDUCTING RESEARCH PRACTICAL STEPS

Research projects have three phases.

The first phase, as you are discovering in this course, is protocol development: the methods and plan of research must be precisely developed and described before research begins.

The next phase is project execution, the steps you must take to ensure that the project is carried out on time and on budget, and that the information you collect is of high quality and will allow you to meet your objectives. The is called the research plan.

The final phase is interpreting and reporting the results.

7.1 Project Execution: your protocol has been accepted for funding – what now?

7.1.1 The research plan

The research plan should begin by listing all the tasks to be undertaken in the study. Responsibility for carrying out each of the tasks should then be assigned to an individual member of the research team. The personnel needs for carrying out the study are determined by estimating the time required and the level of training necessary to carry out the tasks.

A calendar for carrying out the activities is then developed, including all key phases of the research including:

- recruiting personnel;
- training;
- piloting techniques and procedures (if necessary);

- recruiting participation;
- doing the measurements;
- collating the data;
- analysing the results;
- preparing the reports to the participants and community and;
- preparing scientific reports of the results.

When the number and type of personnel has been determined, the responsibilities of each member of the research team should be written down. The responsibilities of the principal investigator include:

- project management;
- quality assurance;
- public relations;
- ethics;
- analysis and reporting of the results of the study to the participants, authorities and the scientific community.

7.1.2 Controlling the quality of procedures and techniques

The procedures of research must follow preciselythose outlined in the protocol.

The precise indicators for evaluating the quality of information and measurements need to be specified and the technique for recording and interpreting them defined.

A number of steps may be taken to ensure that the quality of the research is maintained at a very high level

- pilot testing of untried methods
  Procedures must be tested in advance to ensure that they will work as planned. For example, if your protocol depends on recruiting 10 new patients every day, you should check (either through existing records or with a small pilot test) that this will be feasible. Similarly, all new equipment must be tested and questionnaires pilot tested to ensure that respondents are able to complete them as you expect. Many investigations start with a funded pilot phase, before the main study protocol is developed.
- procedure manuals
  - for each step in collecting the data;
  - for keeping track of the information after it has been collected.

  Manuals of procedures must be developed describing exactly how techniques are to be carried out. In some instances, such instructions already exist; in others, they must

be developed for the study. In addition, a careful plan must be developed and recorded for the management of data. Special attention must be paid to tracking of data obtained, ensuring quality control of data, checking for errors and ensuring secure storage of the results .

- staff training
  Staff who will be collecting the information must be trained for the study. Even for staff with specific clinical expertise (e.g. clinicians, respiratory technicians, radiologists) the investigators need to ensure that these people understand the research objectives. For example, the procedures for eliciting a clinical history are different from those for completing a research questionnaire; technicians who are used to dealing with very ill patients in hospital may not realise the need for different research procedures when dealing with a mostly healthy study population.
- periodic comparison of data quality over centres, testers, etc.
  When more than one person carries out the measurements or when information is being collected at more than one centre, it is essential to make comparisons of the results obtained by the various technicians and by the same technician on different occasions. This ensures that there is close correlation of the results (inter and intra-observer comparisons), or at least that you have a measure of the differences across centres or technicians.
- identify who is responsible for what
  One final way to ensure high quality of data collected, especially for large projects, is to give responsibility directly to smaller research teams or to individuals and to recognise them for this responsibility. Many epidemiological research projects generate a large amount of data, more than can be described in one scientific publication or report. It is often useful to decide in advance, what reports (or what sections of one large report) will be produced from the project and to share the responsibility for these reports among the members of the research team.

7.1.3 Ensuring efficiency

The following measures will help to ensure that the study is carried out within the time frame specified and within the budget allocated.

- written schedule or timeline;
- periodic progress meetings or reports;
- identify which people are responsible for ensuring that the schedule and budget are met;
- share budget information with all staff this need not include staff salary information, but should include the duration of time each staff member is assigned to the project;

- for multi-centre or multi-research group projects, consider a separate budget for each team or centre (but remember to keep funds in reserve for centralised data analysis costs).

7.1.4 Controlling the scope of the project

It is not uncommon for researchers to find themselves wanting to change some aspect of the protocol in the middle of a research project. Reasons for this may include:

- the methods do not appear to be working;

Example:

> You are not able to recruit the patients you had hoped to recruit; the instrument you bought to monitor exhaled breath is unreliable and breaks down too often; the study subjects don't understand how to respond to many of the questions on the questionnaire; etc.

- some scientific event has occurred (a new research paper published, a new drug on the market…) which makes you question your hypothesis or your objectives;
- the results appear to be suggesting that the original hypothesis was wrong (or, that the hypothesis was correct) .

The appropriate response to these situations is to temporarily stop the project and convene a meeting of your research team to consider the following options:

- stop the studyaltogether (and do not continue until you have a new protocol, a new timeline, and a new budget).
  This may well be the correct choice in the case of the first two situations described above. This is also the main reason for pilot testing your methods before you undertake your project.
- decide to continue with the existing protocol
  This is probably the correct choice for the last situation described above (i.e. you think your results are going in one way or another). It is important to remember that, as with people, information tends to appear in groups or clusters.

Example:

> For no apparent reason, you may find that most of the study participants who have asthma show up during the first one or two months of your study, and that for the remaining 6 months, you seldom encounter a subject with asthmatic symptoms. If you stopped your study after the first 2 months you could easily come to a false conclusion about the prevalence of asthma.

Surprisingly, this kind of clustering of data can occur even when you are extracting results from existing databases (e.g. a tuberculosis registry in your region).

> Changing the protocol mid-stream is not an acceptable option. Why not?

Almost every research project is a compromise between scarce resources and study power (i.e. how large a population is being studied). As you have learned in this course, every study protocol includes a sample size/power calculation. If you change the protocol, you will end up with 2 small studies (each one with a slightly different protocol), instead of the one larger study you planned. These 2 smaller studies will have less power to address your objectives. Given the current trend of funding organisations not to fund studies that are larger than they need to be, it is highly probable that your new smaller study will not be capable of addressing your study hypotheses, no matter how carefully you collect the remainder of the data.

Clinical trials often have an interim analysis point with a "stopping rule" built in to their original protocol. This means that the study is designed (with sufficient statistical power) to allow for interim data analysis at a given point in time. If an effect of treatment is seen (positive or negative) that is very strong at this point, the study may be stopped (provided the "stopping rule" criteria are met). Note that even in this circumstance, the decision is simply to stop the study, not to change the protocol and carry on.

7.2 Managing the information collected

When you first collect information for the purposes of a research study, it is extremely unlikely to be organised in such a way that you can immediately see the "results". As you have learned to this point, your data may take many forms.

Some examples include:

- information extracted from clinical records and recorded on paper forms;
- photocopies of death certificates;
- computer files with lung function test results for a series of patients;
- hand completed questionnaires;
- transcribed hospital or clinic records.

> Experienced investigators recognise that collecting the data for a research project takes about $^1/_2$ of the total time required to complete a research project. The steps involved in checking, organising, analysing, and reporting the results generally take at least as much time as does collecting the data in the first place.

Steps in making raw data into "analysable" data.

7.2.1 Checking original data collection forms and records

The first step in data management is to review the raw data records for completeness and accuracy.

Initial checking of paper records

- small projects : check and correct every paper record BEFORE its information is entered into a computer;
- large projects: check a random subset of records, look for common errors, if possible check and correct all records for the places where errors were commonly made.

General rule for correcting raw data

Make changes to the raw data only for obvious errors, but don't make changes, which require a value judgement. For example: in data which was extracted from clinical files; change date of diagnosis if you can verify that an error was made; don't change date of diagnosis just because it seems unlikely.

7.2.2 Coding data

Some raw data can be entered directly into computer files (e.g. date of birth, yes/no responses to questionnaires, number of cigarettes smoked per day, measured size of tuberculin skin test response). For other types of information (e.g. names of drugs, causes of death, clinical diagnoses, job title), it may be easier to sort or classify the information into manageable groups and assign code numbers to each group. The code, rather than the name of the drug (disease, job, etc), is then entered into the computer.

The standard approach is to create a written codebook, which provides a clear set of standard rules for turning the text information into numeric codes. Coding systems may be very complex (e.g. several pages to describe treatment protocols - each one with its own code) or very simple (e.g. "1" for females; "2" for males).

ALWAYS HAVE MORE THAN ONE COPY OF YOUR CODEBOOK!

Precoding:

In many instances, it is advisable to choose codes IN ADVANCE of collecting the data. For example, if you are going to collect information about drugs prescribed, decide in advance how you will classify the drug names and what codes you will use. Then provide this coding information to the person who will be collecting the data. This ensures that the interviewer (or person doing the transcriptions) will be sure to record the information in a way that makes coding possible later.

Another option is to place the codes (and the coding rule) directly on the data entry form.

Some choices regarding coding:

1. record only the code, not the text
   - advantage: reduces one step (less time, less error),
   - disadvantage: cannot verify coding later.
2. record the text, but enter only the code in the computer
   - advantage: reduces one step (less time),
   - disadvantage: need to go back to paper record to verify coding.
3. enter the text directly into the computer and code later, with the computer
   - advantage: reduced one step (less error), can verify coding easily,
   - disadvantage: computerised coding hampered by non-standard spelling, style etc.

In the past, because of computer storage limitations, option 3 was not feasible. It has become more feasible to enter text directly into the computer. Coding may still need to be done in a non-automated fashion (i.e. by the researcher or assistant coding directly into the computer as well), but subsequent verification of coding is facilitated.

> As a general rule, the fewer the number of rewriting operations between the raw data and the computer stored data, the fewer the errors.

Checking coding

Coding should be checked (just as you did for raw data). A standard way to check coding is to recode (without knowing the original codes) a random sample of the data. If an unacceptable number of differences are found, all the data should be recoded and discrepancies reviewed. This is especially important when coding choices require judgement (e.g. diagnoses, causes of death, jobs, appearance of abnormalities on chest radiographs).

7.2.3 Entering data into computer files

Data can be entered into computer files in a number of different ways:

- direct data entry by computer as it is collected (e.g. a computerised questionnaire);
- keypunching coded forms;
- optical scanning or bar codes.

Records and variables

A single data record is all the information you have collected about 1 person (assuming your unit of observation is an individual person) – like a ROW in a spreadsheet.

A single variable is all the information – for every person in your study - you have collected about a particular characteristic – e.g. age, diagnosis, height, response to question about current coughing, etc. – like a COLUMN in a spreadsheet.

Some computerised data entry programs can do some automated error checking (e.g. Epi-Info). The computer cannot find all errors, but you can set up the data entry form to check for out of range values or values which do not make sense.

7.2.4 Checking computer files

After the data has been entered into computer files, it must be checked again.

a) checking for out of range and not valid values
   For variables that are categories (e.g. yes/no, coded variables): have the computer produce a frequency distribution for each variable (i.e. a list of how often every value is present) and examine the results. For variables that are continuous measures (e.g. age, height, FEV_1): have the computer produce a list of the minimum and maximum for each variable, and examine the results.

b) checking for non-sensible values: (e.g. if the person does not have a tuberculosis diagnosis, there should not be values recorded for treatment variables).
   Have the computer produce a cross-tabulation of two variables. Look for logically inconsistent results (e.g. a person coded as having not completed high school, but also coded as being a physician; a person aged 12 years with 3 children; a smoker with no information recorded for duration of smoking, etc).

c) checking for results which seem unlikely
   Examine the frequency distribution (or average and standard deviation) for each variable and ask yourself if the results make sense. For example, if you have conducted a survey of all residents of a particular region you should check the distribution of age and sex. If it does not match up with what you know about the region, this might signal an important error in the way the data was entered. Or, if you have studied nurses and find that 80% are listed as male, you might want to check to see if one or more of the coders reversed the coding.

> "Every experienced investigator knows that even the most meticulous data collection efforts suffer from errors that are detectable during careful editing. If editing is planned as a routine part of handling the data, such errors need not cause serious problems. If editing is ignored, however, problems can result."
>
> from Rothman KJ, Greeenland S. Modern Epidemiology, Lippencott-Raven. 1998

7.2.5 Handling missing data

No matter how carefully you have collected and checked your data there will still be some missing data. **You must decide what to do about this.** Why is this?

Example:

> Assume you have conducted a small study (100 patients) to investigate knowledge and attitudes of patients in relation to the effectiveness of tuberculosis treatment. In your study population you have 12 patients whobelong to a particular social grouping and 10 of these have multi-drug resistant disease. You want to test if belonging to this social grouping is related to multi-drug resistance or if this apparent association is due to other factors (such as education, HIV status, age). However, for 8 of these 12 patients (in this social grouping), HIV status information is missing.

When you analyse your results, every time you do any analysis using both this factor (i.e. belonging to the social grouping) and HIV status, you will only have 4 of these patients included. When you do an analysis with this factor alone, you will have all 12 of these patients included. It is almost certain that your results will be confusing and difficult to interpret, if you have not checked for missing values.

Types of missing information:

- missing "at random"

This does not necessary mean truly random, but means that the missing data is unrelated to other important factors in your study.

Example:

> Missing lung function results on a small number of people because the spirometer broke one day is not random. It is related to a specific testing day. As long as there is nothing about that day which was important to the study, these results are as if they were missing "at random".

- missing "systematically"

This is the most common type of missing data. This means that the missing data are related to some other important factor in the study.

Example:

> If you are studying causes of death and using death records collected by the state, you may find that the records are less complete the further back you go in time. This could mean that you have missing birth date information on the oldest subjects

in your study. If the disease is related to age (or to period of time in history), then you will have a bias in your study associated with the fact that many of the subjects born in an earlier time period will be excluded (for at least some of your analyses).

How to manage missing or obviously incorrect information

1. Deletion

The most common approach in health research is to delete (from the data that you analyse) every study participant with missing data. If you choose this option, you must consider that you now have a modified study population. You can decide to delete only those subjects with missing information for the most important variables in your study, but you must be careful when you analyse results using other variables.

2. Substitution

a) substitute the mean value:

If you can demonstrate that the missing information is truly "missing at random", you can substitute the mean value from all other subjects, for the missing value.

Example:

You are studying an adult working population and for 2% of your subjects no birth date is recorded. First, you check to be certain that this is not related to anything else in your study. Once you are certain of this, you simply assign each of these people the average birth year of your study population.

This option is more often chosen for social science research – e.g. for studies in which very large numbers of people respond to survey questionnaires.

b) substitute an estimated value (with caution!)

Sometimes you can estimate the missing or incorrect value with some accuracy.

Example:

In a study of air pollution and mortality in Mexico City, researchers found a large number of deaths registered in official records with dates that differed from the dates indicated in hospital records. They had two choices. They could either assign a date of death (choosing one source as "more reliable", or taking the average data between the two) or they could delete these subjects from their study. In that study, because they were

interested in the relationship between day of death and air pollution levels just prior to that day, they had to delete the subjects. However, if they had only been interested in describing the mortality patterns on a monthly or yearly average basis, assigning a date of death, based on the other information in the data record would have been the appropriate choice.

CHAPTER 8

INTERPRETING RESULTS

8.1 Ensuring that the results are valid

The value of research depends on the extent to which any inferences drawn from a study, especially generalisations extrapolated to populations beyond the study sample, are warranted. In addition to any inherent limitations of the study methods, questions about the representativeness of the study sample, and the nature of the population from which it is drawn, there are several other problems that can jeopardise the usefulness of a study. In this section we identify some of these problems and discuss how to avoid them or how to manage those that are unavoidable. Among the most important are problems of error and bias.

8.1.1 Definitions

Error: two types of error that may occur in epidemiological studies:

Random error is a capricious inaccuracy in a measurement, which can be due to chance, to imprecision of the testing method or to careless technique in making measurements.

Sytematic error is a consistently incorrect measurement in one direction that may be due to a faulty instrument, wrong technique of measurement, or inappropriate study design.

Bias: this leads to consistent differences between the recorded and "true" values of a variable. Bias can arise at any stage in the collection, analysis, interpretation, publication or review of data. It may result in:

- a mistaken estimate of the frequency of the condition under study;
- incorrect assignment of determinants for the development of disease;
- false assessment of the relationship between a disease and associated determinants.

Reduction in the risk of error and bias requires a clear understanding of their sources and appropriate rigour in the design and conduct of studies. Even so, they cannot always be avoided though it may be possible to compensate for bias (but not error) and/or measure its consequences by appropriate methods of analysis.

The analysis of epidemiological study data entails the calculation and interpretation of rates or average values and comparison of differences in rates or average values among groups. The calculation of rates and average values in study samples are vulnerable to error and bias which can invalidate comparisons between groups and result in misleading conclusions. Errors can occur in the way the study groups are sampled and in the way that the study variables are measured. These will be discussed in detail in this section.

8.1.2 Error and bias in sampling study populations (selection biases)

Inadequate sampling methods

The target population from which the study sample is chosen must be appropriate to the purpose of the study. If it is not the results of the study may not be generally applicable.

Example:

> If a work site with very low levels of exposure to toxic agents is selected for study, no adverse health effect will be identified in association with theexposure. If there is a "healthy worker effect" (i.e. selection for the job in question is based on symptoms or effects of exposure), no effect may be found.

The frequency of exposure to the suspected agent must be sufficiently high for any adverse effect to be detected and selection for work in the study population should not be on the basis of the presence or absence of symptoms that will obscure any adverse effect of exposure.

Other common problems are:

- Failure to define accurately the population to be investigated or from which cases are to be drawn;
- Failure to investigate all eligible subjects from a target or sample population;
- Deviation from rules of selection leading to subjects being unrepresentative of the target population;

- Omission of "hard to find" persons;
- High refusal rate (refusers may include more or fewer of exposed or affected individuals as compared with accepters depending on incentives to participate or not, such as ease of access to treatment or threat to employment). It is essential to trace and enlist participation of all sampled persons;
- High "drop-out" rate, especially if this is correlated with risk of exposure to a suspected agent or an intervention;
- Replacement of refusers, untraceable selected persons or "drop-outs". This is only acceptable if such persons were sampled in error due, for example to the use of an out-of-date list from which the sample was drawn. Volunteers are never acceptable for this purpose;
- Persons initially classified as exposed change their exposure status after commencement of the study ("contamination").

Avoiding error and bias in sample selection.

These errors can be minimised or avoided by:

- use of up-to-date and complete population registers;
- encouraging full participation of all selected subjects by avoidance of inconvenience or discomfort;
- tracing and persuading non-responders to participate fully;
- checking the similarity of the attributes of responders and non-responders to ensure that there is no systematic difference on key variables such as age, gender, marital status, and occupation.

Procedures in analysis to reduce the effect of error and bias due to losses from the study population:

Allowance can sometimes be made at the stage of data analysis for bias due to non-responders and those "lost-to-follow-up" by one of the following methods:

- exclusion of such persons from both numerator and denominator;
- inclusion of all participants on a "time at risk" basis;
- inclusion of all "lost" persons for half the "time at risk" on the assumption that they would have participated on average for half the total duration of the study;
- calculation of separate risk ratios, one assuming that all "lost" persons developed the disease or had the worst outcome and the second assuming vice versa. This will establish a range within which the "true" result might lie.

8.1.3 Error or bias in measurement (information biases)

Sources or error and bias in the measurement or collection of data can be classified under:

- subject variation;
- observer variation;
- limitations of technical methods.

Subject variation results from differences in observations on the same subject on different occasions and may arise because of:

- physiological changes
  Example: lung function in asthma patients;
- factors that affect responses to questions, including accuracy of recollection of past events, motivation, mood, reaction to environment and rapport with interviewer;
- changes induced by being aware of being studied (Hawthorne effect).

Observer variation results from differences in observations on the same subject by the same observer on different occasions (intra-observer error) or by different observers on the same occasion (inter-observer error) and may arise because of:

- awareness of the hypothesis under investigation
  Example: smoking habits in relation to respiratory symptoms;
- errors in execution of the test or phrasing of questions
  Example: an interviewer suggests the "acceptable" response by the method of phrasing a question;
- carelessness or lack of skill or experience of observers
  Example: omission of some questions in a questionnaire or failure to obtain maximum effort in the performance of spirometry;
- bias in execution of tests
  Example: "terminal digit preference" in reading results of tuberculin skin testing.

Technical error may result from failure of the measuring instrument to give accurate or correct results because:

- the test is inappropriate for the purpose
  Example: the use of chest radiographs for the designation of activity status in tuberculosis;
- the instrument is unreliable or inaccurate with the result that the measurements are not repeatable or do not correlate with the severity of the condition being measured
  Example: tests of "small airways function" for the investigation of asthma;
- there are faults in the test system
  Example: a leak in a spirometer or a faulty batch of tuberculin.

- the method of data collection adopted favours obtaining information on one exposure or outcome rather than others
  Example: dust measurements in mines record dust levels averaged over the full work shift, but do not reflect the frequency of peak exposures;

Example:

> If careful standardisation of measurements (such as spirometry) is not carried out, the study results may not be valid. A classic example is the study of Vermont granite workers in which the spirometer had an undetected leak on one of the survey occasions leading to a false conclusion of decline in lung function over time. Note that if the spirometer was reading falsely low by the same amount throughout the study, comparisons between subgroups in the study and comparisons over time would be valid, but comparisons with external populations would be invalid.

Random error may obscure a real difference between groups leading to a false conclusion of lack of effect while systematic error (bias) may lead to an apparent difference which in reality does not exist.

Avoiding error and bias in information collected

The following will help to avoid some of the above errors:

- diagnostic criteria must be clear and rigidly observed, even at the risk of missing a few cases;
- classification of severity or clinical grade should, if possible, be quantitative and cover a full range of possibilities;
- all subjects must be observed and tests carried out under similar conditions and they should be comfortable;
- questions should be kept as simple as possible; check questions may be used to test for consistency of response; (e.g. "do you become breathless on exertion?" and "are you ever short of breath when hurrying on the level?")
- the number of observers should be kept to a minimum; after training, they should be tested for inter and intra-observer variability by repeat observations on the same subject by different observers and by the same observer on different occasions;
- subjects and observers should be "blind" to the hypothesis being tested in order to avoid the risk that personal expectations might influence and so bias the results;
- tests used should be appropriate for the diagnosis of the condition being studied, should be accurate, reproducible and acceptable to the subject;
- test equipment should be robust, simple, reliable and easy to use;

- test methods and questionnaires should be standardised and regular quality assurance procedures applied to ensure that there is no "drift" in the precision of the results.

8.1.4 Assessment of the error inherent in a test or instrument

The power of a test to deliver useful information, given its intrinsic liability to error, can be expressed in several different indices.

Discrimination is the ability of a test to separate those with a disease (or attribute) from those without or to place subjects accurately on a scale of severity (for example, peak flow measurement for asthma, sputum smear microscopy for tuberculosis).

Reproducibility (repeatability or precision) is a measure of the consistency with which a test or question produces the same result or answer in the same subject under similar conditions on successive occasions. It is assessed by:

- replication of tests or questions by the same or different observers using the same instrument on the same group of subjects (or set of specimens);
- comparison of the test system in use with a different instrument or system;
- use of check questions (questions of a similar nature which should produce comparable answers – for example, age and date of birth);
- random allocation of subjects to interviewers for repeat interviews.

A measure of reproducibility is:

Per cent agreement = (number of agreed results/total test positive) x 100

Validity (accuracy) is a measure of the capacity of a test to give true results, that is, correctly to detect the presence or absence of a condition or correctly place a subject on a scale of measurement.

It has two components:

- sensitivity, the ability of a test to identify correctly those with a disease;
- specificitythe ability of a test to identify correctly those without disease.

When analysing continuous variables, the decision on where to place the cut off point between positive and negative values usually involves a trade off between sensitivity and specificity.

Predictive value of a test is a measure of its ability correctly to predict a positive or negative result. It is assessed for both positive and negative results:

Positive predictive value = test positive cases / all cases with disease

Negative predictive value = test negative cases / all non-cases of disease

Predictive value will increase with increased prevalence of the disease, given constant sensitivity and specificity. It will also increase with increased specificity of a test given constant prevalence and sensitivity.

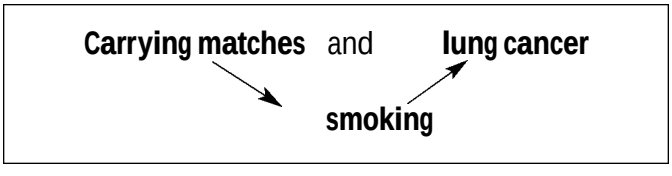
8.1.5 Confounding (a special kind of selection bias)

The interpretations of associations seen in study results may be affected by external variables that are related to both a determinant (risk factor) and the outcome under study - a so-called confounding relationship.

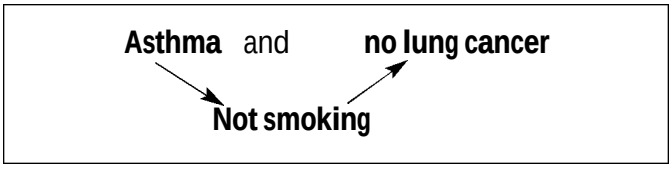
Definition

A positive or negative correlation between a risk factor and a disease that arises because the factor is independently associated with a cause of the disease without itself being a cause.

E.g. Spurious positive: carrying matches causes lung cancer



E.g. Spurious negative: asthma protects against lung cancer



Controlling confounding

Control in design

- randomisation: if sample is large, this ensures potential confounders are equally distributed in all comparison groups;
- restriction: limit entrance to a study to persons within specified homogeneous categories of potential confounders (e.g., confine study to persons from one age group or ethnic origin);
- matching: ensures potential confounders are distributed in identical proportions among study groups (e.g. by matching on smoking habits in both of the above examples). NB There is a danger inherent in overmatching, which will mean that the role of any risk factor that is linked to a matched variable cannot be evaluated.

Control in analysis

- Stratified analysis

Calculate risk estimates for association within strata of confounding variable; an example of controlling confounding by stratification is included in a subsequent section on statistical analysis.

Example:

> If we wished to compare death rates in populations of very different age structure, we could do this by first dividing the populations into age groups (strata), calculating the death rates within each group and comparing the twopopulations within each group (stratum).

- Standardisation

An alternative strategy to control for confounding by age is to calculate a summary value that takes account of the differences in the age structures of the two populations. This is called age standardisation. Although in theory, standardisation can be done for factors other than age, in practice, most standardisation is performed to control for confounding by age.

Indirect standardisation

In this method, the death rates for a particular condition in given strata (by age, gender, etc.) of a general (or standard) population are applied to the study population to obtain the number of deaths expected in the study population (according to the experience of the general or standard population). These are then compared with the number observed, either within each stratum or overall and a proportion (the Standardised Mortality Ratio) is obtained.

Direct standardisation

In this method, the death rates for given strata (for example, age groups) in the study population are applied to a "standard" population age structure. In this method, two populations with very different age structures can then be compared, "adjusting" for the difference in age structure by calculating the number of deaths expected in each study group, from age-specific death rates applied to the same "standard" population.

See the companion text for practical exercises on bias, standardisation, and confounding.

8.2 Making sense of the results - Data Analysis

Doing research is like being a sculptor. One starts with a plain block of marble. Inside this block is a marvellous figure. The task of the sculptor is to "uncover" the image inside this block.

In the same way, the researcher deals with a plain "block" (the mass of information – usually a set of numbers collected and transcribed). There are patterns inside this block that must be uncovered. Just as failing to use the correct tools or to show care in executing the sculpture will fail to uncover the image, so the knowledge that exists will not be revealed if the researcher is not careful in analysing this "block" of information.

See the companion text for practical exercises on understanding the role of statistics in research.

Some useful references for an overview of statistical analysis are:

McNamee R, Cockcroft A. Statistical and epidemiological reviewing. Occup Environ Med. 1994; 51: 721

Greenhalgh T. How to read a paper. (a series of articles from the British Medical Journal; 1997: 315):

- The Medline database. p 180-3
- Getting your bearings (deciding what the paper is about). p 243-6
- Assessing the methodological quality of published papers. p 305-8
- Statistics for the non-statistician. I: Different types of data need different statistical tests. p 364-6
- Statistics for the non-statistician. II: "Significant" relations and their pitfalls. p 422-5
- Papers that report drug trials. p 480-3
- Papers that report diagnostic or screening tests. p 540-3
- Papers that tell you what things cost (economic analyses). p 596-9
- Papers that summarise other papers (systematic reviews and meta-analyses). p 672-5
- Papers that go beyond numbers (qualitative research). p 740-3

Motulsky H. Intuitive Biostatistics. New York: Oxford University Press, 1995.

8.2.1 Getting acquainted with the data

The first step in organising data is to become familiar with it. Just because you have spent a lot of time planning for, and collecting, the information, you are unlikely to know exactly what you have collected until you make a special effort to examine and summarise the data.

Do not move too quickly to analysis of the results, without first examining the data.

Step 1 – examine the variables

Prepare a series of tables showing the distribution for each variable for which data was collected. Review the distribution of each variable – for checking purposes and to determine the shape of the distribution of each variable.

> **TIP:** Make a table for each type of variable (i.e. one for all the "yes/no" or discrete variables; another for all the discrete variable with a small number of response categories; and another for all the continuous variables).

Step 2 – summarise the important information

Discrete (or categorical) variables:

- calculate the rate (or proportion, percentage) of the population with certain characteristics (frequency distribution);
- can be a graph (bar / pie) or table.

Continuous variables:

- make a frequency histogram;
- listing of minimum, maximum, mean (or median), and measure of variability (e.g. standard deviation).

EPI-INFO note:

- to obtain a frequency histogram in Epi-Info for a continuous variable, it is necessary to categorise the variable (into a new variable)
- e.g.:

```
Define age group _____
Recode age to age group by 5
HISTOGRAM age group
```

Step 3 – compare without statistical testing

Discrete variables:

- simple "side by side" comparison of rates (or proportions);
- cross-tabulations;
- rate differences (subtract one proportion from another);
- rate ratios or odds ratios (divide one proportion by another).

Continuous variables:

- "side by side" comparison of mean (or median) values;
- overlay two frequency histograms.

See the companion text for practical exercises on describing study results without statistical analyses.

8.2.2 Approaches to statistical analysis

In order to carry out statistical tests on the data, you must first identify:

- what type of data or measurements you have (discrete or continuous);
- what comparisons or descriptions you want to make.

The most difficult task in statistical analysis is to decide which comparisons to make. Once you have decided how to display your results and which comparisons you want to make, you are now ready to apply statistical tests to these comparisons.

REVIEW: Ask yourself:

1. What was the study question?
2. Is this simply a descriptive study, with no comparisons?
3. What main comparisons are needed to answer the study question (review the hypotheses)?
4. Will you be comparing 2 or more groups (or subgroups)?
5. Will you be comparing 2 or more measures in one group only?
6. Are the factors being studied expressed as a proportion (or rate) or as values measured on a continuous scale?

8.2.3 Descriptive Statistics for describing results, no comparisons

a) discrete data: calculate a rate and the 95% confidence interval around the rate

e.g. in a sample of 2000 children, 123 cases of asthma were found

prevalence rate: 123/2000 = 6.15%

95% confidence interval: 5.1% - 7.5%

b) continuous data: calculate the average value, and the 95% confidence interval around it

e.g. recording age in a sample of 165 adults

average age: 42.7 (standard deviation = 18.5)

95% CI = 42.5 – 42.9

EPI-INFO note:

- to obtain a 95% CI for a rate, the command is: FREQ variable name /C
- Epi-Info does not appear to have a command that gives the 95% CI for a mean (or average value), so here is the formula:

  95% CI = mean ± {1.96 x (std.dev / n-1)}

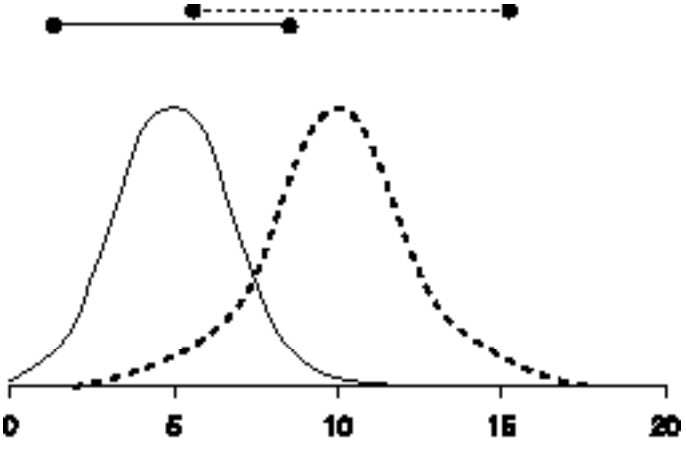
What is a 95% Confidence Interval?

When you select a sample of people to study (from a target population), the result (e.g. asthma prevalence rate) is accurate for the sample. However, since the objective is to understand "truth" in the target population (not just the sample), it is important to know whether the result from the study reflects the true prevalence rate in the target population.

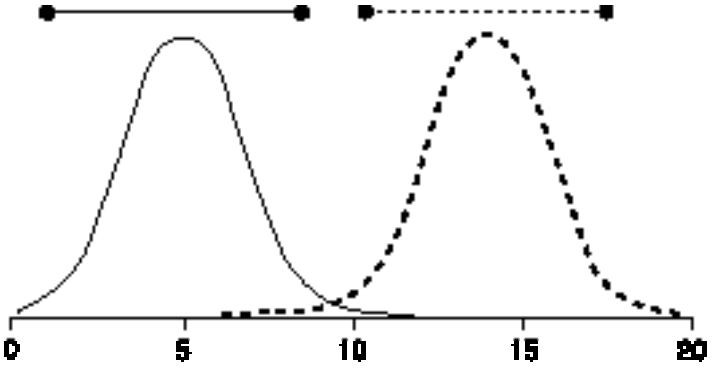
A confidence interval gives an indication of the range in which the true prevalence rate in the target population lies. It is based on the observed rate (in your study), the sample size, and a formula derived from probability theory.

To understand what a 95% confidence interval means, imagine carrying out the same study 100 times in the same target population, using the same methods and the same sample size, but with different study subjects selected each time. If you did this and calculated the prevalence rate for asthma each time, you would get 100 slightly different prevalence rates. Most of the rates would be quite close to each other, but some would be quite different (simply because of chance, or random variation).

The 95% confidence interval is the range in which 95% of 100 results would lie. The narrower the confidence interval, the more confident you can be that your result (in your one study) is a good reflection of the true rate in the target population.



However, if 100 separate samples were taken from another (more different) population and compared to the sample population designated by the solid line above, results may look as follows, with 95% confidence intervals that do not overlap.



8.2.4 Statistical tests for comparing discrete variables between groups

a) comparing 2 groups (or 2 variables) – both discrete, with 2 categories each

Examples of discrete variable with 2 categories:

- Gender (M/F);
- Crowding (yes/no);
- Appropriate case management (yes/no);
- Educational intervention received (yes/no);
- Exposed to high concentrations of irritant gases at work (yes/no);
- Tuberculin reactivity (yes/no);
- Asthma (yes/no).

e.g. to answer the study question: Is there a relationship between asthma (yes/no) and occupational exposure to irritant gases (yes/no)

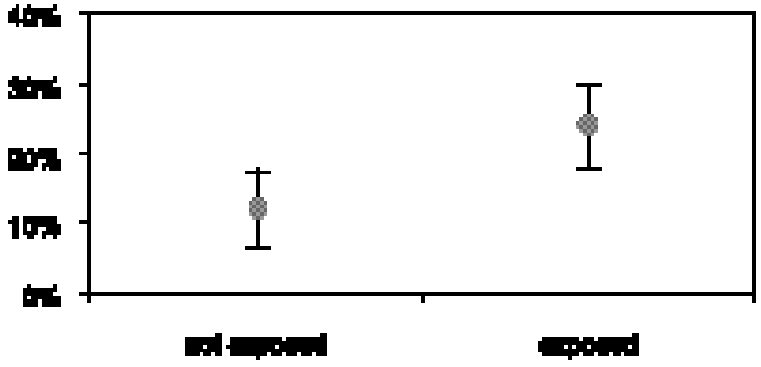
Possible approaches:

(i) calculate 2 rates and calculate the 95% confidence intervals around each rate (and compare these visually)

Example:

> Prevalence of asthma symptoms among workers exposed to high peaks of irritant gases:
> **48 / 199 = 24.1%, 95% CI: 17.1 – 30.6**
>
> prevalence of asthma symptoms among workers not exposed:
> **16 / 130 = 12.3%, 95% CI: 6.9 – 19.3**

Prevalence of Asthma Symptoms



This visual approach gives a good indication of both the extent of the difference in symptom prevalence between the two groups (the size of the effect) and the likelihood that the ratesare truly different. If the 95% CI's overlap a great deal, then you are unlikely to conclude that the prevalence rates in the groups are different.

To calculate the actual statistical probability, use the chi-square test.

(ii) chi-square test

To test the statistical probability that the two subgroups (i.e. workers exposed to irritant gases and those not exposed) actually have the same underlying rate of asthma symptoms, use the chi-square test. This test examines the probability that the two different rates are really from the same underlying "family" or distribution of rates (but by chance, appear to be different).

Chi-square test (comparing 48/199 and 16/130): p=0.008

What does the p-value mean?

The interpretation of this result is that the probability that the two sub-groups actually have the same underlying rate of asthma symptoms (i.e. come from the same underlying "family" of rates) is about 8 in 1000.

EPI-INFO note:

- to obtain the chi-square test result for comparing two rates, the commands are:

```
SET PERCENTS=ON
TABLES variable 1 name   variable 2 name
```

where variable 1 is the determinant and variable 2 is the outcome, or disease

These same data can be displayed in a different way (a 2 by 2 table):

	Asthma symptoms	No asthma symptoms	total
Exposed	48	151	199
Not exposed	16	114	130

You can also use the EIP-INFO program called STATCALC (in which you enter the data as shown above) to calculate the chi-square directly.

If you have a small study and end up with less than 5 persons in one or more of the cells in the 2 by 2 table, use a version of the chi-square test that takes this into account (e.g. Fishers exact test, or "Yates corrected" chi-square). EPI-INFO gives the results for these tests as well as the usual chi-square test. The values are very close to those for the usual chi-square.

(iii) calculate two rates, the ratio of the two rates (the relative risk), and then calculate the 95% confidence interval of the ratio

Using the same example, comparing the prevalence of asthma symptoms among workers exposed, and not exposed to high peaks of irritant gases, compare the two prevalence rates by making a ratio:

$$\text{Prevalence rate ratio} = \frac{24.1\%}{12.3\%} = 1.96$$

A confidence interval can also be calculated for the ratio:

In this example, the 95% CI = 1.2 – 3.3

EPI-INFO note:

- to obtain the rate ratio and its 95% confidence interval, use the same commands as for obtaining the chi-square statistic:

  SET PERCENTS=ON
  TABLES variable 1 name variable 2 name

where variable 1 is the determinant and variable 2 is the outcome, or disease

How to interpret a rate ratio and its 95% Confidence Interval:

Recall that if the two prevalence rates were about the same, the ratio of the rates would equal (or be close to) 1.0. Therefore, if we are interested in knowing if the symptom prevalence rates are not equal, then we want to know whether or not the rate ratio is not equal to 1.0 (i.e. is either greater than or less than 1).

In this example, the rate ratio is approximately 2 (actually 1.96), indicating that the symptom rate is about double in the exposed group compared to the non-exposed group. (We call this value, i.e. 1.96, the "point estimate").

The 95% confidence interval tells us that, if we were to sample the entire population again 100 times, and compare asthma symptom rates, 95 times out of 100, the rate ratio would be greater than 1.2 and less than 3.3. Conversely, this tells us that there is considerably less than a 5% probability (or $p < 0.05$) that the rate ratio is equal to 1.0.

The chi-square test and the confidence interval for the rate ratio tell us the same thing with respect to "statistical significance". The advantage of the confidence interval approach is that it also indicates the magnitude of the difference between the groups, something the chi-square statistic does not.

SUMMARY:

When interpreting rate ratios, we are interested in:

- the magnitude of the ratio itself; and
- whether or not the 95% CI includes (or excludes) 1.0.

b) comparing > 2 groups or subgroups – both discrete, or comparing groups with > 2 categories

Examples of variables with 3 or more categories:

- smoking status: never, former, current;
- severity of asthma: mild, moderate, and severe;
- residential groups: large urban, medium or small town, rural;
- ethnic group.

(i) calculate rates for each group and calculate the 95% CI around each rate (then compare). This is exactly the same as described for 2 rates.

(ii) chi-square test

This test is also the same as for 2 rates, but it compares all 3 (or more) rates. It determines if they (jointly) come from the same underlying "family" or if one or more might be from a different "family" of numbers.

Example:

	Non-smokers	Former smokers	Current smokers	P-value (chi-square)
Total number studied	259	267	196	
N (%) with airflow obstruction	20 (7.4%)	34 (13.2%)	26 (13.3%)	0.05

(iii) rate ratios

You cannot calculate a single rate ratio when there are more than 2 groups, therefore, you must rely on the chi-square test alone to compare more than 2 groups.

In some situations, you can calculate a ratio of any two of the 3 or more rates. This should only be done if the chi-square test indicates that the groups are not the same, and if your study plan was to compare 2 specific groups.

For example, if you study hypothesis involves comparing never smokers to current smokers, you could exclude former smokers in the example above, and calculate a relative risk for airflow obstruction and smoking as follows:

Prevalence of obstruction among smokers:	13.3%
Prevalence of obstruction among never-smokers:	7.4%

Relative risk = 1.78; 95% CI: 1.03 – 3.10

Why this restriction?

> With 3 or more groups, there will always be one with the highest rate and one with the lowest rate. If you simply test if these 2 groups are different, the statistical test may APPEAR to support the conclusion that they are. This approach is wrong. You must test FIRST, to see if all groups come from the same underlying population. Only then can you compare pairs of 2 groups (called 2 way comparisons). Only make two way comparisons that are consistent with your study questions.

8.2.5 Statistical tests relevant for continuous measures: compared across 2 or more groups

Tests for measures that are distributed "Normally" (or approximately so).

a) 2 groups or subgroups

(i) calculate the mean values and their 95% confidence intervals (and compare/look for overlap)

Example:

In a randomised controlled trial of the use of a new combined therapy, compare the age of patients in the experimental therapy group to patients in the conventional therapy group:

	Experimental therapy group	Conventional therapy group
Age (years), mean (std dev)	33.5 (17.4)	36.8 (16.2)
95% CI:		
if n=120 in each group	33.2 – 33.8	36.5 – 37.1
95% CI:		
if n= 15 in each group	31.1 – 35.9	34.5 – 39.1

EPI-INFO note:

- to obtain the mean values, use the command:

  MEANS variable 1 name variable 2 name /N
  where variable 1 is the continuous variable and variable 2 is the grouping variable.

  The "/N" part of the command prevents a long listing of all the data being printed.

  This command does NOT give 95% CI's. You can calculate them, using the formula:
  95% CI = mean ± {1.96 x (std.dev / n-1)}

(ii) Analysis of Variance (ANOVA)

z-score test (for large groups)

t-test (for groups with n < 30)

These statistical tests are essentially the same as comparing the confidence intervals to test the hypothesis that the 2 groups come from the same underlying population. The results are given as "p-values". The p-value is the probability that the two groups DO come from the same underlying population (i.e. that the groups are NOT different).

Following the example above:

	Experimental therapy group	Conventional therapy group	p-value*
Age (years), mean (std dev)	33.5 (17.4)	36.8 (16.2)	
95% CI:			
if n=120 in each group	33.2 – 33.8	36.5 – 37.1	< 0.01
95% CI:			
if n= 15 in each group	31.1 – 35.9	34.5 – 39.1	> 0.10

* t-test or ANOVA: probability that groups are from the same population

b) 3 or more groups or subgroups

ANOVA (analysis of variance)

This statistical test is a z-score test to test the hypothesis that the 3 or more groups come from the same underlying population. The same point as made above for rates applies here, i.e. You must test FIRST, to see if all groups come from the same underlying population and only then compare pairs of 2 groups. Only make two way comparisons that are consistent with your study questions.

Tests for measures that are NOT distributed "Normally" (i.e. skewed)

- where you would otherwise use a t-test, use the Mann-Whitney U test (unpaired data);
- where you would otherwise use ANOVA, use the Kruskal-Wallis analysis of variance (by ranks).

EPI-INFO note:

- to obtain ANOVA results, or Mann-Whitney and Kruskal Wallis tests, use the same MEANS command:

  MEANS variable 1 name variable 2 name /N

where variable 1 is the continuous variable and variable 2 is the grouping variable.

8.2.6 Statistical tests relevant for continuous measures: association between 2 measures

Tests for measures which are distributed "Normally" (or approximately so).

a) Looking for association between 2 different measured values (in one group)

i) plot the 2 measures against each other – look at the relationship

EPI-INFO note:

- to plot 2 continuous variables against each other:

  SCATTER variable 1 name variable 2 name

  (variable 1 will be plotted on the "x" axis and variable 2 on the "y" axis)
- if you put a "/R" after the variable names, Epi-Info will draw the "best" STRAIGHT line for the plot

ii) Pearson correlation coefficient ("r")

- the correlation coefficient "r" tells something about the STRAIGHT line association between 2 continuous variables;
- "r" can be either positive, ranging from 0 (meaning no association) and 1 (meaning that the 2 variables are exactly the same), or negative (between 0 and –1) ;
- "r^2": tells what proportion of the total variance is shared between these 2 measures (or roughly, what proportion of the variance in one is accounted for by the other – and vice versa); note, "r^2" always varies between 0 and 1.

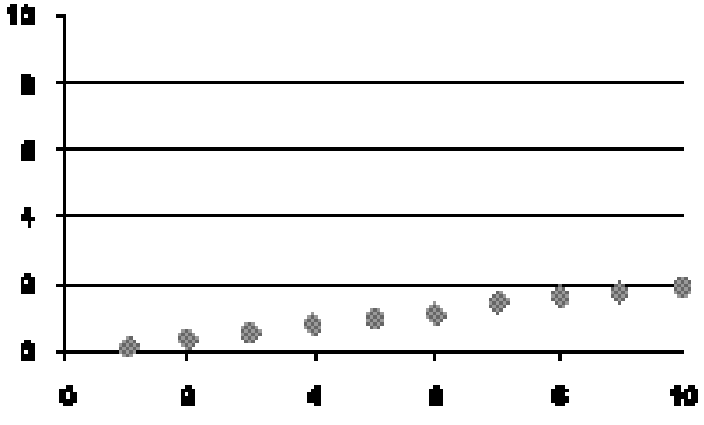
b) Looking at whether one measure explains or predicts the other

- Where the correlation coefficient tells **how closel**y associated two variables are, the linear regression coefficient tells **how much** the "y" variable changes for a given change in the "x" variable;
- mathematically: it is the SLOPE of the best straight line relationship between the variables;
- regression coefficients can take any value (positive or negative);
- the value depends on the units of the two measure being compared – it is simply the change in "y" for every "1-unit" change in "x";
- linear regression coefficient (classed "beta"): looks only at the extent to which one measure explains variance in the other **(and not vice versa)**.

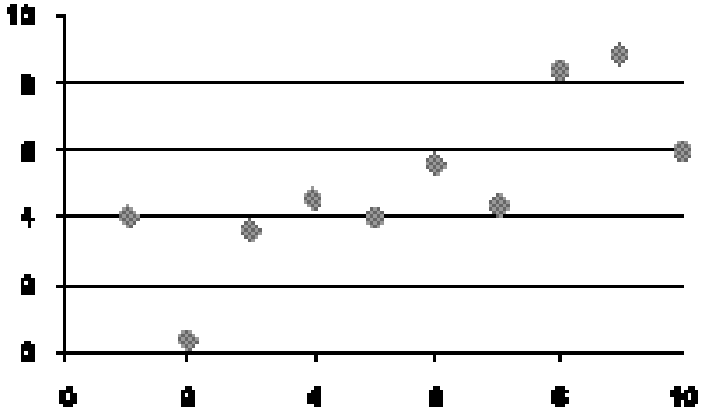
"p-value" of correlation and regression coefficients

The p-value for a correlation coefficient and for a regression coefficient ONLY tells you about the probability that the coefficient is equal to zero. It does not tell you anything about how different from zero the coefficient is.

Example:



the regression coefficient (slope) for this relationship is 0.2 and the p-value is <.001 compared to:



this relationship has a regression coefficient (slope) of 1.0 but a p-value of > 0.05

EPI-INFO note:

- to obtain the correlation coefficient and the regression coefficient for theassociation between 2 continuous variables:

  REGRESS variable 1 name = variable 2 name

where variable 1 is the "outcome" variable, or the variable you want to explain; and variable 2 is the "determinant" or the variable you expect may predict or explain changes in variable 1

- the output shows the "r", the "r^2", the beta coefficient, and the 95% CI for "r^2" and beta

8.2.7 Statistical tests that consider comparisons, while "controlling" for other variables

An unbiased study design requires that the two (or more) groups being studied are similar in all respects except for the factor or factors you are studying. Your research design should try to ensure that this is the case. However, in epidemiology, in practice this is seldom the case.

Therefore, you need to consider how your groups compare (at baseline) and then "control" for differences between your groups.

Example:

> We studied airflow obstruction rates among 2 groups of workers, those exposed to high levels of wood dust (n=174) and those not exposed to wood dust (n=126). The two groups had different age distributions.
>
> **Question:** Is age a potential confounder (of the relationship between airflow obstruction and exposure to wood dust) in this study?
>
> **Answer:** A confounder is a variable related to the health outcome (in this case, airflow obstruction) AND (by chance) also related to the exposure.

Note: by definition, we do not call a factor a confounder if it is related to the exposure because it is part of the pathway between exposure and disease.

Data from the study:

	Age ≤ 40	Age > 40	R.R. (95% CI)
Number studied	137	163	
% with airflow obstruction*	2.9%	16.6%	5.7 (2.4, 13.7)
% exposed to wood dust	70.1%	47.9%	1.7 (1.3, 2.3)

* FEV_1 < 80% predicted

As shown above, in this study age was related to both the health outcome and to the exposure, so it is a potential confounder.

Statistical approaches to controlling for confounding:

a) Stratify

If we simply examine the relationship between airflow obstruction and exposure to wood dust in this study, we obtain the following results:

	Exposed to wood dust	Not exposed to wood dust	R.R. (95% CI)
Number studied	174	126	
% with airflow obstruction*	12.1%	7.9%	1.5 (0.7, 3.1)

* FEV_1 < 80% predicted

The results suggest a slight, but not statistically significant, increase in airflow obstruction among the wood dust exposed group.

However, when we divide the exposure groups into age groups (e.g. < 40; > 40) and compare rates of airflow obstruction in association with exposure within the strata, we obtain:

	Exposed to wood dust	Not exposed to wood dust	R.R. (95% CI)
Number studied	174.8%	126.8%	
Age < 40 years			
% with airflow obstruction	4.1%	0	**** 3.9
Age > 40 years			
% with airflow obstruction	21.8%	11.8%	1.9 (0.9, 3.8)
Mantel-Haenszel			
summary relative risk			2.1 (1.1, 4.2)

*** logit approximation

The Mantel-Haenszel summary relative risk (and MH summary odds ratio – for case-control studies), is a weighted relative risk that tests the common association between the risk factor (exposure to wood dust) and disease (airflow obstruction), after controlling for age group.

In this example, we see that, once differences in age between the exposure groups was taken into account, workers exposed to wood dust were about twice as likely to have airflow obstruction than those not so exposed, and that this difference was statistically significant.

EPI-INFO note:

- to obtain stratified analyses:

  SET PERCENTS = on

  TABLES variable 1 name variable 2 name variable 3 name

where variable is the determinant, or predictor variable, variable 2 is the health outcome, or disease, and variable 3 is the potential confounding variable, or variable you want to control for

Detecting Interaction by stratification

Most statistical software packages that calculate MH summary risk ratios also test for significant interaction (e.g. in Epi-Info this is called “test for homogeneity of the odds, or risk ratio”).

This tests the hypothesis that the association (e.g. the rate ratio or odds ratio) is different in the different strata (i.e. that the relationship between airflow obstruction and wood dust exposure is not the same for young compared to older workers). Another word for interaction is “effect modification”.

In the example above, the relative risks were not exactly the same (i.e. the RR was 1.9 in younger workers and 3.9 in older workers) but the test for "homogeneity" of these odds ratios was not statistically significant.

Therefore, there was **confounding** (because, by bad luck, our groups differed in age), but no significant **interaction** (that is, once we controlled for the age difference, we found greater airflow obstruction in exposed workers from both age groups).

Example:

The following example, from a case-control study of asthma among adults in the general population shows the relationship between asthma and exposure to substances at work suspected of triggering asthma, controlling for smoking status.

	Asthmatics	Non-asthmatics	O.R. (95% CI)
Number studied	173	285	
% smokers	19.6%	27.7%	
% with high risk exposures at work :			
Smokers and non-smokers combined	24.7%	16.4%	1.7 (1.1, 2.6)
% with high risk exposures at work:			
non-smokers only	25.9%	13.6%	2.1 (1.1, 4.2)
Smokers only	17.7%	24.0%	0.7 (0.2, 1.8)
MH summary OR			**1.7 (1.0, 2.7)**

Breslow-Day test for homogeneity of the Odds Ratios (i.e. test for interaction) p=0.04

In this example, the "crude" OR for the relationship between asthma and work exposure (i.e. the odds ratio without any stratification) was the same as the MH summary odds ratio, taking smoking into account. This is because, although smoking was related to asthma (i.e. there were more non-asthmatic smokers), smoking was not related to exposure (the proportion of smokers was about the same among exposed (27%) and non-exposed (26%).

Therefore, smoking did not confound the overall relationship between asthma and work exposure.

BUT, stratification revealed that there was an important difference in the relationship between exposure and asthma for smokers compared to smokers, i.e. that smoking modified the relationship. This INTERACTION (or effect modification) would not have been apparent without stratifying the data.

b) Standardise – this is discussed above, in the section on confounding

c) Modelling – using multiple regression

One of the limitations of stratification to control for other variables is that it is necessary to categorise your variables. Also, stratifying for one variable usually works well, but if you want to "control" for more than one variable, you quickly approach tiny numbers in the strata.

Multiple regression models allow for the effect (on the outcome) of more than one risk factor to be examined together. Essentially, these models test the effect of one risk factor while controlling for all other factors in the model.

What does multiple regression analysis do?

Basically, multiple regression analysis does the same as simple regression analysis (described earlier) – i.e. it tells how much the "y" variable (the health outcome measure) changes for a given change in any one of the "x" variables (the potential predictor variables in the regression model), while controlling for all other variables in the model.

A regression model is simply an equation with the following form:

Outcome variable ("y") = intercept (a) + (coefficient 1, β_1 X predictor variable x_1)
+ (coefficient 2, β_2 X predictor variable x_2)
+ (coefficient 3, β_3 X predictor variable x_3) etc.

Although multiple regression modelling is relatively easy with many modern statistical analysis computer programs, strategies for deciding what should be in the model and for interpreting the results are more difficult. Researchers not familiar with modelling should seek the assistance of colleagues with experience and training in these methods.

Multiple regression overcomes some of the problems with stratification, but introduces new problems associated with trying to "control" for many variables at the same time. Because you are still limited by the size of your study population, you must be very careful not to include too many variables, and not to include variables that are highly correlated with each other.

For example, if you are studying the risk of tuberculosis infection you may wish to take into account socio-economic status, geographic region, household crowding, presence of an infected contact, and income level. However, the fact that these potential predictor variables are all associated with each other will make the results of multiple regression modelling difficult to interpret.

- for a continuous outcome (e.g. FEV_1) – use linear regression

Just as with simple linear regression, mathematically the linear regression coefficient is theSLOPE of the best **straight line relationship** between the outcome and the predictor variable, controlling for all other variables.

- for a binary outcome (e.g. disease: yes/no) – use logistic regression

Mathematically, the regression coefficient (β) from a logistic regression model can be used to calculate the ODDS RATIO for the association between the health outcome and a "1 unit" increase in the predictor variable, using the formula: $OR = e^{(\beta)}$

Some statistical software packages do the calculation for you and provide the 95% CI for the Odds Ratio.

EPI-INFO note:

- to carry out multiple LINEAR regression:

  REGRESS variable 1 name = variable 2 name variable 3 name… etc.

where variable 1 is the health outcome, or the variable you want to explain – and variables 2,3, and so on are the "determinants" or the variable you expect may predict or explain changes in variable 1

the output shows the r^2 (for the whole model), the β coefficient and its 95%CI for each predictor variable.

TIP: if you want to include discrete predictor variables in a regression model, code them so the "1" represents "yes" and "0" represents "no". The regression coefficient is (mathematically) the numeric increase in the outcome variable for a "1-unit"' increase in the predictor variable. This will allow you to interpret the regression coefficient as the change in "y" or the outcome variable when comparing the "yes" to the "no" category of the predictor variable.

Example:

In our study of wood dust exposed workers, the following results were obtained from a multiple regression model to examine the association between FEV_1 (as % predicted) and age (in years), current smoking (recorded as 0=no, 1=yes), asthma (0=no, 1=yes); and wood dust exposure (0=no, 1=yes).

REGRESS

FEV_1 (%predicted) = age smoking asthma wooddust

	Coefficient	95% CI	p
Intercept	111.6	107.1, 116.1	0.0001
Age (years)	- 0.31	-0.41, -0.22	0.0001
Current smoking (0=no, 1=yes)	- 2.12	-4.27, 0.03	0.05
Asthma (0=no, 1=yes)	- 5.84	-9.09, -2.58	0.005
Wood dust exposures (0=no,1=yes)	- 2.98	-5.07, -0.88	0.005

These results indicate that:

- for every 1 year increase in age, FEV_1 (%predicted) decreased by 0.3%;
- among current smokers (compared to non-smokers), FEV_1 (%predicted) was 2.1% lower;
- among asthmatics (compared to non-asthmatics), FEV_1 (% predicted) was 5.8% lower;
- among workers exposed to wood dust (compared to those not exposed) FEV_1 (% predicted) was 2.9% lower.

What does the 95% CI and p-value mean here?

If there is NO relationship between the outcome and any one predictor variable, the coefficient would be 0 (or close to it), indicating no change in the outcome associated with that factor. It follows that if the 95% CI does not include 0, there is a significant association between the outcome and the determinant.

Therefore, in regression the p-value is an indication of the probability that the coefficient is zero.

8.2.8 Other issues in statistical analyses

a. paired vs. unpaired?

Paired analyses should be done if you have measured something 2 times on the same person (or group) or when study subjects are individually matched to comparison subjects.

Example:

- paired t-test (in place of a regular, unpaired t-test);
- Wilcoxon matched pairs analysis (in place of a Mann-Whitney U test– for skewed data).

b) outliers

During the descriptive analysis stage, you may have discovered extreme or idiosyncratic values in your data. These could be either an error or a truly correct (but unusual) value.

First, consider the possibility that the data represents a study subject from a different population than the one you planned to study.

Example:

In a study of the relationship between lung function and workplace exposure, we encountered a non-smoking, non-asthmatic subject with extremely low lung function values. It turned out that this subject had only 1 lung).

If the value is in error – or represents a truly different study group, you may choose to delete that subject from the study. However, if you cannot justify deleting the subject, you must be alert to the possibility that the extreme values may exert a strong influence on some of your comparisons.

c. multiple comparisons

If a statistical test indicates a p-value of 0.05, this means that the result you found could have arisen by chance alone 5 times out of 100 (or 1 time in 20).

So… if you make 20 comparisons in a study, there is a good chance that at least 1 such comparison will have a p-value of 0.05, by chance alone. This is the "problem of multiple comparisons".

d. interpreting "negative" results

When study results are not "statistically significant" it is worthwhile to consider the possibility that this is the result of a study too small to test the hypotheses. You can never answer this question definitively, but you should look at the upper confidence interval (e.g. of the rate ratio) and consider if this upper confidence limit is in the clinically relevant range.

e. parametric vs. non-parametric tests

"Parametric" refers to statistical tests that assume that distributions are roughly "Normal"and that the amount of variability in the data is roughly the same at all levels. The descriptive data analysis stage is necessary in order to determine which types of tests you need to use.

f. results based on very large sample sizes

It is not uncommon for comparisons or associations to be "statistically significant" in very large studies, even if the comparisons are not interesting from a biological or health point of view.

g. correlation/causation

The presence of a statistically significant association between a risk factor and a disease outcome in a research study indicates simply that the factors are related. Other evidence should be considered before concluding that the association is "causal". These factors have been discussed elsewhere in these course notes.

8.3 Reporting the results

Research is not complete until the results are presented to the partners of the research, are made available for independent review by external experts and are submitted for publication in the peer-reviewed scientific literature.

8.3.1 Preparing a report of the research

The research should be described in a formal report to the partners in the research (discussed in a subsequent section). This report normally follows the outline of the research protocol and makes use of much of the materials prepared for the protocol. This includes the background of scientific knowledge, the description of the population and methods, the presentation of the results, along with a discussion of their significance and a list of recommendations that flow from these results.

8.3.2 Writing a scientific paper

There is no one particular way to write a scientific paper. However, there are some general rules that should be followed. A scientific paper contains a title and the following sections: abstract, introduction, materials (subjects) and methods, results, discussion, and references.

Before starting to write the paper you should identify the journal most appropriate for publication of your study. You should then review the instructions for authors published in the journal. These instructions should be followed precisely (deviation from them will likely result in rejection of the paper). You should take care to read some papers published in the journal on the same topic as your paper to get an impression of the style of the papers preferred in the journal.

Title

The title should be concise and catch the reader's interest to your paper. Turning the title into a question may sometimes be a good idea. For instance, if you have examined the association of socio-economic status to treatment failure in tuberculosis the title could be:

"The relationship between socio-economic status and treatment failure in patients with tuberculosis" or the title could be; "Is socio-economic status related to treatment failure in tuberculosis?"

Abstract

Most journals state in their instructions to authors how the abstract should be written, and how many words it should contain. This should be followed in detail. The abstract is often written after you have finished the rest of the paper. It is important that the content of the abstract is in accordance with that of the text of the paper.

Introduction

You may start the introduction by identifying the subject of your paper in general and its importance. You should then focus more precisely on your topic. To update the reader you should then briefly review what has previously been published on this topic and identify what is not known from these previous studies.

If you, for instance, want to examine to what extent private general practitioners in Pakistan stick to the guidelines for TB management, you may start the introduction as follows:

"Every year there are about 8 million new cases of tuberculosis world-wide. To ensure correct treatment and follow-up of these patients, guidelines for management of tuberculosis have been published. Knowledge as to what extent these guidelines are put into practice is a prerequisite for intervention programmes to ensure that they are followed. In countries A, B and C, X-Y % of the physician was aware of these guidelines. In Pakistan no such data is available."

In the following part of the introduction you should state your hypothesis and finally the objectives of the study. Not more than 2 or 3 objectives should be stated.

The length of the introduction should be about $^1/_2$ of a page to one page with double-spaced text in 12-point type.

Material and methods

Start describing the study design and the selection of your study population. If a case control study is performed, describe from where your cases and controls were sampled. The response rates and any dropout rates in both cases and controls should be described. In case of an intervention study similar data should be presented in the intervention groups. Sometimes it may be useful to present a figure of the study design.

You should define the outcome or dependent variable and the determinants or independent variables. Furthermore, you should describe how these variables were measured. This description should be given in sufficient detail to allow other researchers to evaluate or reproduce the tests or experiments. For methods that have been published previously, provide only a brief description of the method and state the references where they are described. If drugs have been given, you should include the generic name, dosage and route of administration.

If you have used a questionnaire, a brief description of it should be given, including number of questions and what topics they deal with. You should then present the wording of the questions used in this particular paper, including the alternatives for answering. If you have used a standardised questionnaire, it is enough to give a reference to where the wording of the questions was published.

The statistical handling of the data should be provided. Describe the statistical methods used for each analysis. If the tests are not well known, give references to them. You should also provide sample size estimations. The presentation of the sample size estimation is often done in a standardised way. If you, for instance, want to examine the effect of an intervention on failure rate in TB treatment you may present the power calculation as follows:

"A reduction in treatment failure rate from 10 to 5% was considered clinically important. Based on a power of 80% and a significant level of 5% we would need 431 subjects in the intervention group and control group, respectively. With an anticipated response rate of 70% a total of 616 subjects were needed in each group."

All work involving studies on human beings must receive approval from an ethics review committee, according to the Helsinki Declaration. It should be confirmed in the material and method section of the paper that such approval from the ethics review committee was obtained.

Results

You should include only the important results, that is, results that help answer the questions of the study. In general most of the results should be given in tables or figures. The text should emphasise the most important observations, or the most important parts of the figures or tables.

The results section should start with a brief description of the study sample. You should then describe the relationship between the outcome variable and the determinants based on univariate analyses. Finally you may describe these relationships in multivariate analyses. The latter analyses imply that the independent associations of each of the determinants to the outcome variables are examined.

The text of the results section should be brief; it should not be longer than 2 pages with double-sized text in 12-point type. You should not include more than 6-8 figures and/or tables. The main message may often be more easily presented in a figure than in a table. The disadvantage of a figure compared with a table is that you can usually present fewer details.

Discussion

The discussion should start with a brief summary of the main results of your study. You should then discuss the methods used, their advantages and limitations compared to other methods that could have been used. You should comment on possible implications on the results from choosing these particular methods. For instance, if you have examined patients with tuberculosis, you may discuss your method of diagnosing tuberculosis compared to other methods that could have been used.

Furthermore you should discuss the validity of the study, both the external and internal validity. You should then discuss the main findings of the study, and explain contradictory or unexpected results and compare the findings of your study with those of previous studies.

In order to attract the interest of the editor, you should emphasise the new and important aspects of the study. It may also be wise to discuss the implications of your findings and state possible topics for future research. The latter aspect is particularly important if you already have started research in one of these topics.

Finally, in the discussion section, you should state the conclusions of the study. It is important that the conclusions are in accordance with the objectives stated at the end of the introduction section.

It is usually not possible to include among the authors of the paper all those who have made the study possible to conduct. It is a custom to acknowledge those who contributed to making the study possible but who were not included in the list of authors in a section on "Acknowledgements" after the discussion section.

References

The references should follow the format recommended by the "Instructions to Authors"of the journal to which you intend to submit the article. You should take care to present the references in exactly the way the instructions direct. They are usually numbered consecutively in the order in which they first appear in the text. Most journals prefer not more than 30 to 40 references.

When you have completed writing the paper, it is often useful to send it to someone outside your research group to obtain critical comments. When your paper has been returned from the editor, you must consider carefully the comments by the reviewers and reply specifically and precisely to the comments made. If the paper was rejected for publication, you should rewrite the paper, taking into account the critical comments made by the reviewers, and then submit it to another journal.

CHAPTER 9

OTHER ISSUES IN RESEARCH

The technical aspects of design and conduct of research often preoccupy us, but there are other important issues to be considered.

9.1 Who owns the research?

Research is an activity undertaken by a number of players working together: these include the investigators, their staff, the institutions they represent, those who provide the finances, the participants, their communities and the institutions in which the research is carried out.

9.1.1 Structuring responsibility

Since research is a partnership, ideally this should be explicitly affirmed in the formation of a "steering committee" comprising representatives of all the partners. Usually the investigators and the sponsors of the research meet regularly (or communicate) to set out the terms of reference, conditions and time-table of the research. Other important partners such as representatives of the community in which the research is to be carried out and the regulatory or administrative agencies responsible for their health are less often included.

The steering committee should begin to operate from the conception of the research and should take an active part in the planning process. It should also assist in carrying it out. This is vital in ensuring that the results of the research are communicated and those policy recommendations arising from the research are implemented.

9.1.2 Addressing perceived needs

In the process of carrying out research, the perceptions of those being studied (the community) concerning the health-related state and determinants under investigation need

to be taken into consideration. This may be partly achieved by having representatives of the community on the steering committee. It may also require public meetings with members of the community to address their questions, to gain their co-operation and to explain the aims and procedures of the research. The community may have important questions which have not been included in the original protocol but which might be easily addressed within the context of the study. Sometimes investigators assume that the community will have nothing to add to an essentially intellectual enterprise such as research. Many investigators who have learned much through their collaboration with the communities they have studied can attest that this view is mistaken.

9.1.3 Ensuring follow up of the results

If the research has been carefully planned, a steering committee has been formed and the community engaged, follow up of the research becomes much easier.

Investigators must keep in mind that the aim of research is to create knowledge but the purpose of knowledge is action for health. To achieve this action, the results of the research and clear recommendations for action must be communicated through appropriate channels (see below) to the community and to those responsible for the health services in the community.

Sometimes research investigations may uncover previously unknown disease or medical information of importance to the health of the participants. A well-defined plan of follow up to deal with such instances, never breaching the confidentiality of individual participants, must be in place.

9.2 Getting support for research

Brundtland, formerly Chair of the World Commission on Environment and Development and now Director General of the World Health Organization, has stated that "Developing countries must build up their own basis for research. Only they will be able to establish the diagnosis and implement the cure. The international community must assist the process".

9.2.1 Local sources

If one considers all the resources (both human and material) expended on research in low-income countries, unquestionably the countries themselves contribute the lions' share.

Relevant research that is also cost-effective almost always originates at the bedside of the "patient" (or, in the case of public health activities, the program). Research and practice must always go hand in hand. This is the reason why the Commission on Health Research for Development has proposed (and many donor agencies have accepted) that a portion (2-5%)

of the budget of all public health programs must be set aside for research. This is the most important source of support for research because it is more likely to be targeted to the research relevant to practice in the community.

A second source of financial support for research at the local level is central government. This support is usually directed through the official channels of the country such as universities, research institutions. To access such support, investigators usually must have collaboration with such institutions or groups. For clinicians and program managers, this should be a fruitful way to carry out research. Should there be support available from both the programs themselves and the special research funds, the two can sometimes be joined leading to highly productive research which takes its hypothesis from the field but harnesses research expertise such as epidemiology or statistics to carry it out.

A third source of financial support for research at the local level is the pharmaceutical industry. If the topic is of particular interest to the firm, research support may very well be made available. Much of the recent research in low-income countries into the distribution and determinants of asthma has been funded in this way.

A final source of local funding is the development agencies, both governmental (through the embassies) and non-governmental. Frequently the development departments of local embassies have limited (but often sufficient) funds for support of humanitarian and community action. Service organisations, such as the Lions or Rotary Clubs, may provide support to research at the local level. If a good rapport can be developed with the officials of such programs, and if the research can be shown to have clear practical relevance to the situation in the country, funds might be obtained.

9.2.2 International Sources

International agencies, which support research in low-income countries, are few in number but fund a substantial amount of such research. The agencies may be multilateral (the United Nations system and the European Union) or bilateral (Cooperation agreements between industrialised and low-income countries).

Multilateral agencies

The United Nations system is the most active agency in this regard. Most of the assistance for health research from the United Nations is provided through the World Health Organization (WHO). This budget increased greatly during the 1980s to over 50 million dollars annually. Most of this budget (over two-thirds) was spent on disease prevention and control. It was managed mainly through two agencies of the WHO, the Human Reproduction Program and the Tropical Diseases Research Program. The latter program has spent a large amount of money on research into malaria, schistosomiasis, filariasis,

trypanosomiasis, leishmaniasis and leprosy. Interestingly, although lung diseases are the single most frequent cause of death in the world in small children and young adults, relatively little money has been committed to research in this area. Future priorities for activities in WHO will include tobacco and its prevention and control and this may be a fruitful area for obtaining research support in the near future.

The European Union has a very large budget for health research and has specific structures to promote research collaboration with low-income countries. This research, however, almost invariably takes the form of international collaborative research and usually requires the collaboration of several research centres in Europe in order to be eligible for funding.

Bilateral agencies

Traditionally, there were only two agencies specifically mandated to provide support for health research in low-income countries. These were SAREC, a Swedish Government agency (which now works in conjunction with the European Union) and IDRC, a Canadian Government agency whose specific mandate is research on development. Traditionally, it has supported a great deal of health research but more recently has less of a focus on specific health research. The current emphasis in this agency is on Health Services and Systems research.

Other bilateral co-operation agencies do fund research, although that is not their specific mandate. The Japan International Co-operation Agency funds a number of projects in collaboration with their citizens in a variety of low-income countries. Other agencies (for example, in the USA and UK) frequently channel their support through national research institutions in their own countries.

Foundations

A number of international foundations have provided research support. The main foci of research in these foundations have been infectious and tropical diseases (one third) and epidemiology, policy and management (one-third). Major donors to health research in low-income countries have been:

- The Aga Khan Foundation, Switzerland (primary health care management),
- Carnegie Corporation USA (human resources development),
- Edna McConnell Clark Foundation USA (tropical diseases research),
- Ford Foundation USA (reproductive health),
- MacArthur Foundation USA (tropical diseases, women's health),
- Pew Charitable Trusts USA (health policy and management),
- Rockefeller Foundation USA (population, neglected diseases),
- Sasakawa Memorial Health Foundation Japan (leprosy),
- Wellcome Trust UK (medical and veterinary research).

9.3 Ethical issues in research

The ethics of medical research, especially that carried out in low-income countries, have been hotly debated in recent years. The issues were summarised in the "Declaration of Helsinki"[8]. Those specifically relating to research in low-income countries were addressed by the Council for International Organisations of Medical Sciences in a publication in 1982[9].

9.3.1 Basic principles

Biomedical research involving human subjects

The basic principles guiding biomedical research involving human subjects are as follows:

Technical aspects

- research must conform to accepted scientific principles and be based on laboratory and animal experimentation as well as a sound knowledge of the scientific literature;
- research procedures must be fully described in a protocol and reviewed by an independent research ethics committee;
- those carrying out research must be scientifically qualified and supervised by a competent medical person;

Liability

- responsibility for the well-being of the participants must always rest with the medically qualified individual and not the subject;
- the importance of the objective of the research must be in proportion to any inherent risk associated with the research procedures;
- a clear statement must be made of predictable risks in relation to foreseeable benefits to the subject and to society;
- the interests of the subject must always prevail over the interests of science and society;
- the integrity (including the privacy) of the subject must be ensured;
- research can only be carried out if the hazards are predictable and benefits must outweigh potential hazards;

(8) Adoped by the 18th World Medical Assembly, Helsinki Finland in 1964 and revised by the 29th World Medical Assembly, Tokyo Japan in 1975.

(9) Council for International Organizations of Medical Sciences. Human experimentation and medical ethics. Geneva: Bankowski Z, Howard-Jones N, ed. 1982, pp 505.

Credibility

- accuracy and lack of bias in reporting the results of the research must be assured;
- researchers must always declare fully any potential conflict of interest;
- research not complying with the basic principles should not be accepted for publication.

Communication and consent

- subjects must be fully and adequately informed of aims, methods, benefits and hazards;
- subjects must be informed that they are free to abstain or withdraw from the research at any time without prejudice;
- subject must give consent, based on full information, prior to participation;
- consent must be free and given to someone to whom the subject is not dependent;
- consent must be given by a legal guardian where legal incompetence prevents the subject from providing it.

Procedures of the protocol

- the research protocol must always contain a discussion of ethical considerations and should declare its adherence to the principles of the Helsinki Declaration.

Clinical research (research combined with medical care)

Additional principles apply to clinical research.

- new tools for diagnosis or treatment may be used by a doctor if they offer hope of saving life, re-establishing health or alleviating suffering;
- the benefits, hazards and discomfort of new tools must be weighed against those of current methods;
- all participants must be assured of access to the best methods of diagnosis and treatment;
- refusal to participate must not interfere with the delivery or quality of care;
- where consent of the participant is not requested, the reasons for not doing so must be explicitly stated and a justification presented;
- the research must promise potential diagnostic or therapeutic value to the patient/ community.

Non-clinical biomedical research (not involving care)

- the responsible doctor must remain the protector of the life and health of the research participants;
- the research must be discontinued immediately if it has caused any harm or is perceived to be at risk of harming any participant.

9.3.2 Principles relevant to sponsored research in low-income countries

All biomedical research, wherever it is conducted and under whatever auspices, must follow the principles outlined above. Some research carried out in low-income countries is sponsored from outside the country, often from donors based in industrialised countries. Such research involves ethical considerations in addition to those outlined above. Issues of particular concern are:

- the investigation may serve external rather than local interests;
- external collaborators may lack insight into local mores, customs and legal systems;
- local disillusionment may result from lack of long-term commitment to the subjects and their community;
- compensation for injury may be jeopardised by lack of accountability.

These issues demand an explicit commitment to the host country and its scientific community through research, which is informed by local priorities, providing service to the community or giving training to local scientists and practitioners. Mechanisms must be in place to respect local mores, customs and legal systems and to assure compensation for injury sustained in the course of the study. The difficulties foreseen and the steps whereby they are addressed must be clearly spelled out in the research protocol.

The varying interests involved in collaborative research and, in particular in sponsored research in low-income countries, must be explicitly acknowledged and discussed. Investigators and authorities in low-income countries should not agree to such research if there are no clear and stated benefits for the subjects, the local investigators and / or the community. The protocol of such studies should clearly specify the interests of each of the collaborating partners and evaluate the "risks" and "benefits" of the collaboration itself.

Ethics review procedures

- all research must be approved by both scientific (technical) review and by ethical review;
- the ethical principles to be followed in all such research should conform to those outlined in the "Revised Helsinki Declaration" (outlined above);
- independent ethical review must be undertaken in the community in which the research is carried out;
- the standards of the ethical review must be at least as rigorous as those in the communities from which each of the investigators come;

Informed consent

- attention must be paid to any indication that participants may be unable to comprehend information about the objectives, risks and benefits of the research and to give informed consent. In such cases explicit mechanisms, acceptable to the local ethics committee, must be proposed to overcome the problem;
- consent of the leaders of communities should be obtained for the research proposed, especially if the members of the community are particularly poor or are illiterate;
- even where community leaders give consent to a research protocol, individuals must be given the right to withdraw from the study, without individual prejudice.

Legal liability

- participants in research who suffer injury as a result of their participation must be eligible for compensation;
- eligibility for compensation for injury resulting from the research cannot be waived by either the participant or a guardian;
- providing a mechanism for compensation is the responsibility of the sponsoring institution, organisation or person;
- one mechanism by which compensation can be assured is by an insurance scheme.

9.3.3 Appropriate structures for ethics review

The first requirement to be met by investigators before they submit a protocol for ethics review is assurance that the technical (scientific) content of the protocol is of an acceptable standard (usually by independent peer review).

Ethics review is essential for all biomedical research carried out on human subjects. An Ethics Review Committee normally undertakes it. The Committee's role is to satisfy themselves that all proposed interventions are acceptably safe for human subjects and to ensure that all other ethical considerations (outlined above) are adequately addressed. Membership of an Ethics Review Committee should include a variety of health professionals with relevant competence and lay persons representing the community's cultural and moral values. Members serving on a review panel must declare the absence of conflict of interest in relation to the individuals submitting the protocol and in relation to the subject matter to be studied before the ethics review is undertaken.

Where procedures exist for ethics review, these must be followed. If there is no established local procedure, it is the responsibility of the investigators to ensure that appropriate mechanisms are set up and the requirements of ethics review are met. No research should be commenced before ethics review and approval.

Ethics review must be based on material submitted by the investigators, including the protocol itself and usually a form specially designed to address ethical issues. The investigators must provide:

Overall

- a clear statement of objectives;
- a summary of current state of knowledge;
- ustification for undertaking the research;

Technical aspects

- a clear statement of the qualifications of the investigators;
- description of all interventions including dosages, duration and any known potential risks to participants;
- a statistical plan including numbers of subjects to be recruited;

Subject participation

- criteria for admission to and withdrawal from the study;
- means of eliciting informed consent.

Research that does not explicitly meet the criteria of ethics review must not be accepted for publication in scientific journals.

Conception et réalisation :
Les Editions de l'Aulne
Paris