Operational research for malaria control

Tutor's Guide

World Health Organization
HIV/AIDS, Tuberculosis and Malaria
Roll Back Malaria

July 2003

Trial Edition
© World Health Organization 2003

All rights reserved.

This health information product is intended for a restricted audience only. It may not be reviewed, abstracted, quoted, reproduced, transmitted, distributed, translated or adapted, in part or in whole, in any form or by any means.

The designations employed and the presentation of the material in this health information product do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers’ products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

The World Health Organization does not warrant that the information contained in this health information product is complete and correct and shall not be liable for any damages incurred as a result of its use.
Table of contents

Foreword .....................................................................................................................................................3
Acknowledgements .....................................................................................................................................4
Introduction .................................................................................................................................................4

Learning Units
1. Methodology I: research question and rationale ..................................................................................4
2. Methodology II: study types and designs .............................................................................................4
3. Pharmacovigilance and adverse drug reactions (ADRs) monitoring ....................................................4
4. Operational research .............................................................................................................................4
5. Data collection and management ..........................................................................................................4
6. Data analysis I: frequencies and defining measures of central tendency..............................................4
7. Data analysis II: the normal distribution, Univariate and Bivariate analysis .........................................4
8. Data Analysis III: Multivariate analysis ..............................................................................................4
9. Test of associations: risk, relative risk, odds ratio and attributable fraction ......................................4
10. Sources of error ....................................................................................................................................4
11. Budget and plan of action .....................................................................................................................4

Annexes
1. WHO budget requirements ..................................................................................................................4
2. Summary of the Nuremberg code ..........................................................................................................4
3. Summary of the declaration of Helsinki ...............................................................................................4
4. Summary of IEF guidelines ..................................................................................................................4

References ..................................................................................................................................................4
Foreword

This module uses a training method based on learning by problem-solving to facilitate the understanding of the operational research in different epidemiological situations. The underlying principle is that learners who are actively involved through a series of group exercises and discussions learn more and better than those who simply sit and listen to a single person talking for long periods of time. The reasoning and deduction required in the module makes this subject extremely suitable for this training method, but the success of the module will depend on your active participation in the training activities proposed. The module is addressed to health personnel responsible for malaria control at national and sub-national levels of the health care system. It requires some basic knowledge of epidemiology and statistics, malaria case management and some aspects of vector control. However, the contents of the module are flexible enough to allow the emphasis to be placed according to the specific training needs. The main objective of this module is to inform professionals of the basics and methods of operational research. Countries need expertise who can undertake routine operational research on certain interventions for policy changes and to monitor trends in morbidity and mortality. This will therefore help health workers operating in different epidemiological and socio-economic circumstances understand the use and methods of operational research for decision making, monitoring and evaluation of malaria control activities.

The module is divided into two parts - Part I the Learner's Guide and Part II the Tutor's Guide. The Learner's Guide covers basic concepts and information together with a series of problems and hints or partial solutions to them. The Tutor's Guide outlines the main points to be learnt, but does not provide definitive and inflexible responses. In this way it is designed to stimulate active learning.

The module has been conceived for group work. The exercises in the Learner's Guide need individual and group works. Group exercises should be carried out in small groups to stimulate discussions and exchange of experience between the participants (who would come from different countries/areas with different experiences), the facilitators and the tutor. The guide can be used for workshops of varying duration between 2-4 days depending upon the time available and the rate at which the exercises proceed. The module can be independently given in a separate course or be customized into a course with other subjects depending on the need of audiences. Certain exercises may be completed at a later date by the participants individually provided they have both the Learner's and Tutor's Guides. The complete module is optimally designed to be accomplished in 24 hours (3 days).
Acknowledgements

The contents of this module has been developed by Dr Andrew Y Kitua, National Institute for Medical Research, Dar es Salaam, Tanzania and Dr Maru Aregawi from the Malaria Control Department, WHO Headquarters, Geneva. This module takes into account various formal and informal contributions from the members of the RBM Technical staff. Mrs Assil Farah has made tremendous effort to edit the document and finalize the layout.

The module is still a trial edition which needs further editing effort and scientific review. It will be field tested in various international training courses before reaching its final stage of development. Authors highly appreciate inputs and useful suggestions from readers (tutors, facilitators and participants) to be incorporated into further editions.
Introduction

The planning and implementation of a malaria control programme must be based upon epidemiological analysis and application of interventions suitable to specific localities or countries. Health workers and all involved need to be equipped with a sound and updated methods on operational research for different purposes at national, district and peripheral levels. This module can be used for in-service training or as part of a basic course on malaria control. The operational relevance of the understanding of how this knowledge should be utilized for the latter case, is that it is recommended to participants have some basic knowledge on epidemiology and statistics, case management and vector control.

For whom is this training module intended?

The module is designed for health professionals involved in malaria control at national, sub-national and district levels who have responsibility for planning, executing malaria control, and monitoring activities in their respective working levels. These include medical officers, medical assistants, public health officers, environmental health officers, parasitologists, and biologists involved in malaria control either with government or NGOs.

Objectives

At the end of the training programme based on this Learner's Guide you should have acquired the skills that will enable you to:

- Understand methods of operational research and different types of studies.
How is the course run?

Tutor

The tutor has overall responsibility for the planning and management of the course and will also introduce each of the learning units, but the tutor will not give formal presentations of this module.

Facilitators

The tutor is assisted by a number of facilitators who will work with you continuously through small group sessions and provide additional information whenever required. They will also assist the moderators in guiding group discussion. Together with the tutor, they are your constant source of information and experience. If you study in small groups but without a facilitator, the tutors must to some extent play the role of the facilitator.

Presentations

Lectures are kept to a minimum and will be replaced by limited introductory remarks by the tutor at the beginning of each subject and short examples to overcome points of common difficulty.

Small Group work

The module is designed for 3 days of training, working mainly in small groups, say 2 or 3 groups of 6 to 9 learners each. It is desirable for each group to have its own room, with at least one of the following: overhead projector, whiteboard, blackboard, flipcharts. For each unit the group selects, among its members, a moderator and a rapporteur by rotation, so that, as far as possible, each learner performs each of those two functions at least once.

The sessions provide good opportunities for you and the other learners to give your opinions, develop your ideas and learn from one another. The learners will usually have different backgrounds, in terms of training and experience, so that they should have much to learn from each other. The exchange of experiences among participants contributes to most of the training material, the Learner's Guide providing a lead for discussions and work. A moderator chosen by the members of each group will lead discussions on the particular subjects proposed in the learning units. At the end of the group work devoted either by the moderator responsible and discussed by all participants and commented on by the tutor. These presentations and discussions are important but are not meant to be formal as working notes. The overall success of this training module will depend on the active participation of all learners in the group exercises and discussions.
This training module

Use of the Learner's Guide

This Learner's Guide consists of instructional materials and problems designed to enable you and your colleagues to achieve the objectives stated earlier. The Guide is divided into Learning Units. Before each session you should read each Unit carefully and make sure you understand it, as the tutor will not be giving a detailed presentation of the material to be learnt. If you are unclear about any part of the Learning Unit you should discuss it with your colleagues in the discussion group, your facilitator and with the tutor, if necessary. Each Learning Unit consists of a series of questions (and hints and partial solutions to some of them) to be worked through as individual or a group. The discussions during small group work and during plenary sessions with the participation of facilitators and tutors will facilitate this process. You must acquire the skills and knowledge contained in one unit before progressing to the next, otherwise you may have difficulty in achieving the objectives of subsequent learning units.

Individually, make maximum effort to read some of the important references and guidelines sited in the document as details are left for further reading. Annexes are given as additional sources for in-depth knowledge.

Use of the Tutor's Guide

During the course, the tutor's guide would be available only to the tutor and facilitators and upon completion of the course/module, all learners would get a copy of the tutor's guide so that they can use the materials for further training and reference.

The module consists, in its present state, of two major learning units addressing. Each unit consists of a Learner’s Guide and a Tutor’s Guide. The Learner’s Guide proposes a series of exercises and offers hints for some of the problems. The Tutor’s Guide gives guidance to the tutor for answers to the exercises.

The module aims at developing an approach, namely the critical analysis of precipitating factors of malaria epidemics, preparedness and responses under different epidemiological situations rather than to convey a body of facts (even though many facts may be conveyed in the process). Most facts and details are referred to relevant guidelines and other resource materials.

No document can, and this module does not, exhaust such a wide and dynamic subject. Malaria epidemics is dynamic issues and the prevention and control methods also evolve over time and so this module does. The module will be successful if it helps the learners understand the mechanics of malaria epidemics in the context of new developments to incorporate better prevention and control approaches. This will help participants continue to update their knowledge as an integral part of their professional activities.

The Learner’s Guide can also be used in conjunction with the Tutor’s Guide, for individual active self-learning.
Evaluation

Evaluation of the learner

The evaluation of individual progress and achievement will be carried out by the tutor, the facilitators and yourself. It will include:

- **Spot tests**
  At regular intervals, a series of "spot tests" will be set out for you to comment on. They are designed to help you and the tutor assess how well you have mastered the skills and developed the competence to carry out your work.

  Correct answers will be supplied after the spot tests and a discussion will take place. This is intended to improve the process of learning and help you to identify those activities in which you need further practice.

- **Multiple-choice quizzes**
  In multiple-choice quizzes, each question is provided with a list of possible answers from which you must select the one you think is correct. At the end of these sessions you will not necessarily be given the correct answer to each question, but the tutor will analyse the results to identify topics that were not clearly understood. The tutor may also tell you where you made mistakes and point out areas where mistakes were made and point out areas where you need to improve.

This part of the evaluation is designed to help you and the tutor to assess how well you understand the course. Multiple-choice tests will take place at the end of the module to assess the achievement of technical competencies by the participants.

1.1 Evaluation by the learner of the training

At the end of the course you will be asked to complete a questionnaire to tell the tutor how you think the training has helped you and how it might be improved. This evaluation will take place at the end of the training period in order to provide as much feedback from the learners as possible. During the course you should also feel completely free to make suggestions for improvements on the part of the tutor and facilitators as well as in the content of the course and the training facilities. This will help your colleagues in a next training course!
Learning Objectives

By the end of this unit you will be able to:

- Understand how to generate your own research questions.
- State clearly the research problem(s) and rationale.
- Understand the Null Hypothesis and its relation to the working hypothesis
- Understand the ethical considerations in health research

This and the other entire learning unit are presented to you in the form of series of questions which should be answered by the trainees individually or in groups. This is not an examination but is designed to help stimulate and guide the learners thinking process and make them learn by doing. The trainer should let the trainees work on the answers to the questions first and allow discussion to defend their work before guiding them to the correct answer or approving their answer as the correct one. The attitude of the trainer should be positive and encouraging always, and not discouraging by exposing trainees ignorance.

Since the trainees will have different experiences and are likely to be at different levels of understanding, they are likely to differ in the speed of answering questions. The tutor must therefore be quick to identify the slow ones and apply supportive strategies, like placing them in groups of individuals with similar learning capacities, depending on the prevailing situation.
Questions and answers:

1.1 Determining the research Question on Malaria

1.1.1 What are the operational gaps (problems) which, if solved, could enhance malaria control activities in your country? List them under health system, preventive and control measures, and community issues.

The answers may be within the following areas:

- Health System – service issues (Drug availability, quality, safety, access, distribution, affordability, skills and capacity of health workers etc); policy issues (policy guidelines, regulations and their applicability);
- Preventive and control measures – availability and use of insecticide treated nets, insecticides, intermittent preventive treatment, environmental management, availability of effective drugs, cost of the drugs etc.
- Community issues – knowledge, attitudes, behaviours and practices, self medication etc

1.1.2 For operational activity mentioned above state the favourable conditions that could enhance the attainment of best results (highest effectiveness) and the possible level of achievement when applied at community level.

Answer:
The favourable conditions are: Its level of effectiveness, affordability by target population, acceptability, level of coverage and perceived benefits by the community. If the operational activity was improving accessibility and use of insecticide treated nets the important conditions include high level of coverage, use and re-treatment.

1.1.3 Use the table below to fill in information regarding your problems raised above.

<table>
<thead>
<tr>
<th>The Operational Gap (Problem)</th>
<th>Its Contribution to Malaria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Morbidity</td>
</tr>
<tr>
<td></td>
<td>Increase</td>
</tr>
<tr>
<td>1 Low accessibility and use of insecticide treated nets</td>
<td>x</td>
</tr>
<tr>
<td>2 Low compliance to malaria treatment regimen</td>
<td></td>
</tr>
<tr>
<td>3 Poor malaria case management</td>
<td></td>
</tr>
<tr>
<td>4 High resistance to current first anti-malarial drugs</td>
<td>x</td>
</tr>
<tr>
<td>5 Poor access to health care services</td>
<td>x</td>
</tr>
<tr>
<td>6 Lack policies to protect vulnerable groups</td>
<td>x</td>
</tr>
</tbody>
</table>
1.1.4  From the above could you list the three priority problems and explain why?

<table>
<thead>
<tr>
<th>Priority</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>High resistance to current first antimalarial drugs</td>
<td>It increases morbidity, mortality and becomes even more costly by increasing use of health services due increase of severe cases</td>
</tr>
<tr>
<td>Low accessibility and use of insecticide treated nets</td>
<td>It increases morbidity, mortality and becomes costly by increasing use of health services</td>
</tr>
<tr>
<td>Lack policies to protect vulnerable groups</td>
<td>Exposes vulnerable groups to increased morbidity, mortality. It also hinders regular use of health services</td>
</tr>
</tbody>
</table>

Note: The priorities may differ depending on the country situation

The listed priorities should have as much supporting evidence as possible and such evidence should focus on:

- The size of the problem locally, regionally and internationally
- The affected population
- Its potential for spread if not checked
- Its effect on the health services (the burden it exerts on the health system)
- Its economical impact on the affected population

Lack of or under-utilization of effective preventive strategies/tools fore the problem.

1.1.5  Write a one to two page summary stating the problem in relation to what literature shows has been done inside and outside your country on the priority problem identified in (1.1.4). Your write-up should justify why there is still need for research on this subject.

Answer:

The summary should focus on:

- Magnitude of the problem
- Affected groups
- Interventions already done on the issue and their level of success
- Methods of research used (how reliable and relevant they were)
- What areas still need research
- How will your research provide additional information for problem solving on this subject
1.2 Literature review

1.2.1 Different sources of literature are available to-date. Which of the following sources are currently available in your institution? Please tick

- library [    ];  Medline [    ];  online journals access [    ]; books [ ]; The Internet [    ];
- Others state ………………………………………………………………

Give four reasons why literature search is important in defining your research question.

Answer:
Literature search helps to:
- provide current knowledge of the problem
- avoid unnecessary repetitions and re-inventing the wheel.
- Shed light on neglected areas
- Define new approaches, knowing the extent to which previous approaches have been successful, and hence identify the new knowledge the researcher intends to bring about

1.2.2 Refer to your work on 1.1.7 and consider whether you have searched sufficiently to justify your question. Do you think that your research will add to current knowledge or provide new approaches to solving the problem?

YES [     ]  NO [     ]

If YES continue with 1.3. If NO, try to make additional search and reformulate your question accordingly.

1.3 Research Hypothesis and the Null Hypothesis

1.3.1 Using the problem you have chosen and justified for research above, write down a working hypothesis for your study.

Answer:
The working hypothesis should include the following elements:
The intended intervention and beneficiaries,
Measurable effect of the intervention,
Expected duration of intervention necessary to produce the intended effect.

For example: improved access and use of insecticide treated nets will reduce infant mortality by 30 percent in district A within 6 years.

Or

Users of insecticide treated nets will have a 30% reduction in infant mortality within 6 years as compared to non users (whose infant mortality level will remain unchanged)
1.3.2 State the Null Hypothesis for your chosen study

**Answer:**
The Null Hypothesis:
Improved access and use of insecticide treated nets will have no effect in the reduction of infant mortality in district A within 6 years and if there is any demonstrable effect it will have occurred by chance.

1.3.3 Discuss and describe the Null Hypothesis. How does it differ from the working hypothesis.

The Null Hypothesis denies the existence of any significant differences between any compared values and stresses that even when such differences are observed, they have occurred by chance.

The working hypothesis is formulated in a way as to reject the null hypothesis. It states that there is a significant cause of the observed differences between compared values and that the difference did not happen by chance.

1.3.4 How do you interpret results where the working hypothesis (alternative hypothesis) is confirmed?

When the working (alternative) Hypothesis is confirmed, we reject the null Hypothesis and accept that there is a true difference between the Hypothesized value and the population mean.

1.3.5 **How do you interpret results where the working hypothesis is not confirmed? (That is the null hypothesis cannot be rejected?)**

When the working hypothesis is not confirmed, we are unable to reject the Null Hypothesis. However it does not mean that we have confirmed that there is no difference between the Hypothetical and population mean or values.

It only means that within the limitations of our study, we were not able to reject the Null Hypothesis.

1.4 Research Objectives

1.4.1 Give two reasons as to why we have to develop research objectives instead of just doing the research?

**Answer:**
In order to better clarify and focus our intentions
In order to have a basis for measuring our achievements at the end of the study
1.4.2 Well defined objectives make the research strong and credible. What do you consider to be the characteristics of good research objectives?

**Answer:**
Specific, Measurable, Achievable, Reliable, and Timely (SMART)

1.4.3 Referring to your working hypothesis, write down the broad objective of your study.

**Answer:**
To reduce infant mortality by 30 percent within a period of six years by improving access and use of insecticide impregnated bed nets in district A.

1.4.4 Write down the Specific Objectives of your study.

**Answer:**
From the broad objective stated above, the specific objectives could be:

1. To increase availability of insecticide impregnated bed nets at retailer shops and MCH clinics to 100 percent within 3 years in district A.
2. To increase the proportion of households with ITN by 60 percent within 3 years in district A.
3. To increase utilization of insecticide impregnated bed nets at household level to 80 percent within 4 years in district A.
4. To ensure that 80 percent of under five children sleeping under ITN within 4 years in district A.
5. To determine reduction of under five mortality by 30 percent at the end of six years of the intervention in district A.

1.4.5 Assess the above specific objectives by ranking them in terms of being: Specific, Measurable, Achievable, Reliable, and Timely (SMART). Use the following table to perform the work, and for each cell give a rank from 1-3 (1=least, 2=Moderate, and 3=Most).

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specific</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
1.5 Ethics in Health Research

Throughout its history, Medical practice, whose primary aim is to prevent individuals from disease or cure them from disease has been concerned with causing harm to the individual patient.

The initial rules for ethical conduct in medicine are found in the Hippocratic corpus, and the Hippocratic oath aims at providing maximal possible benefit to the sick and protect them from harm and injustice.

However, it may be said that it was the Nuremberg Trial in 1947 for crimes against humanity, which triggered global concerns over the well being of human subjects in medical experiments and resulted in the development of the Nuremberg Code. Its central focus was on “Voluntary consent”.

This formed the basis for development future codes of ethical conduct including The Declaration of Helsinki in 1964 and the Council for International Organizations of the Medical Sciences (CIOMS), “International Guidelines for Ethical Review of Epidemiological Studies in 1991”.

Summaries of the Nuremberg Code, the Declaration of Helsinki and the 1991 Ethical Guidelines for Epidemiologist are provided for reference in Annex II, Annex III, and Annex IV.

It will be noted that the current guidelines focus on Autonomy, beneficence, non-maleficence and justice.

1.6 What do you consider as important ethical considerations in research involving human subjects?

**Answer:**

Important ethical considerations in conducting health research (involving human subjects) include:

- Ensuring minimal risk to subjects or groups by carefully involving non invasive procedures and weighing the benefit against the risks.
- Ensuring just distribution of the burden and benefits of the research through better selection of representation from all groups of the society in the study.
- Respect of individuals autonomy
  - Obtaining informed consent
  - Respecting individuals privacy and confidentiality
- Ensuring that individuals and communities will benefit from the results of the study.
- Ensuring that benefits in terms of health standards achieved during study will be maintained (communities or individuals will be maintained (community or individuals will not be deprived of services provided before).
- Ensuring that there are clear responsibilities of the researcher towards the individual and community and these are clearly defined.
- Obligation to inform the subjects or communities of the study results and towards publication.
2.1 Study types and designs

Research studies can be classified into two major categories. These are
- observational studies
- experimental or interventional studies

2.1.1 What are observational studies and which are their four main types?
Answer: Observational studies record happenings under different conditions and try to relate the outcome to conditions under which the affected individuals were subjected. The researcher does not manipulate the status of exposure, but merely records events based on identified exposure status. The four types of observational studies are:

- Case control studies
- Cohort studies
- Case series
- Crosssectional studies (surveys)
2.1.2 Compare and contrast between Case Control Studies and Cohort Studies

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Case Control Studies</th>
<th>Cohort Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directional</td>
<td>Backward or retrospective follow-up</td>
<td>Forward or prospective follow-up</td>
</tr>
<tr>
<td>Exposure</td>
<td>Exposed and non exposed are not yet known</td>
<td>Exposed and non exposed are known</td>
</tr>
<tr>
<td>Disease</td>
<td>Diseased and non diseased are already known</td>
<td>Diseased and non-diseased are not yet known</td>
</tr>
<tr>
<td>Prevalence or risk incidence</td>
<td>Prevalence can be determined</td>
<td>Follow up for risk incidence</td>
</tr>
<tr>
<td>Measures of Comparisons</td>
<td>Odds ratio</td>
<td>Relative risk</td>
</tr>
</tbody>
</table>

2.2 Experimental or interventional studies are studies in which the researcher manipulates one or more of the conditions under which study subjects are exposed. In this way, the researcher is able to show the extent to which the outcome of interest can be modified intentionally. Mention four types of experimental or interventional studies.

**Answer:**
Experimental or interventional studies include:
- Clinical trials – randomised controlled (with self controls or with external controls)
- Clinical trials randomised but not controlled
- Nonrandomised trial
- Community based interventions – randomised and controlled
- Community based interventions – non randomised but controlled.

2.2.1 Describe the advantages of controlled interventions/trials over non controlled trials

**Answer:**
The advantages of controlled trials are that:
- They provide greater evidence for the cause and effects.
- Provide the possibility of controlling for biases through randomization.

2.2.2 Describe the advantage of randomisation

**Answer:**
The advantage of randomization is that it provides the mean for controlling for biases. Proper randomization will ensure that the compared groups are similar in all other ways except the intervention in question.
2.2.3 Which approach best suits the priority research you chose in (2) (that is which will allow to achieve the objectives efficiently). Explain why you stated so.

**Answer:**
Response will depend on the choice. Tutor to provide guidance and lead discussion.

2.3 **Sampling Procedures**

2.3.1 What do you understand by sampling and why is it important to sample subjects?

Sampling is a procedure for selecting candidates for a particular study or intervention in such a way that they represent the populations from which they were drawn. It is important to sample study subjects in order to save time and money.

2.3.2 List the four common sampling procedures of your knowledge

Sampling procedures include:

- Simple random sampling
- Cluster sampling
- Systematic sampling
- Stratified sampling

2.3.3 In a clinical study to test the efficacy of a new malaria vaccine on reducing the incidence of uncomplicated malaria, 1000 children aged three months were recruited. The researcher wanted to make certain that the allocation of the trial vaccine to half of the children and placebo to the other half was balanced in relation to other exposure characteristics. What method could the researcher employ to attain this wish?

**Answer:**
Randomise the children to the vaccine and placebo groups.

2.4 **Sample Size Calculation**

Sample size is the number of subjects selected to represent a given study population.

2.4.1 Why is it important to determine an appropriate sample size?

**Answer:**
It is important to determine the appropriate sample size because we are required to make inferences to the study population based on our findings from the sample. The size should be sufficient therefore to represent the characteristics of interest of the study population.
2.4.2 What is the basic information required for the calculation of a sample size for a descriptive study or population survey such as descriptive social study or cross-sectional survey?

**Answer:**
Descriptive Studies or Population Surveys:
- Size of the original population
- Expected frequency of the disease (factor or problem)
- The worst acceptable results (what margin of error is allowed in %)

2.4.3 What is the basic information required for the calculation of a sample size for a cohort or case control study?

**Answer:**
Calculating the sample size of Cohort and Case control studies requires information on:
- The confidence level (90%, 95%, 99%) which is the probability that the comparable populations differ.
- Power (80%, 90%, 95%), probability that if the two populations differ, the study will detect the difference “significantly”.
- The ratio of unexposed to the exposed
- Expected frequency of disease in the unexposed group.
- The expected relative risk or odds ratio.

2.4.4 You are to determine the sample size for conducting a randomised double blind placebo controlled malaria vaccine trial in a population living under intense and perennial transmission. The population size is 12600 inhabitants and there are estimated 2100 households. Children under five years comprise 30 % of the population and 20% is comprised of children aged 5-15%. Recent statistics show that on average children under the age of five will get three malaria episodes per year, while those above 5 years will get one episode per year. Two neighbouring communities of 15,000 population and similar characteristics are located across the river which forms the northern boundary of this community. The vaccine is intended to reduce malaria morbidity by 60% In children. Infant mortality is a t 120 per thousand live births.

What would be your sample size to give you a 90% confidence for obtaining significant results?. What if you changed the confidence level to 80%?

2.4.5 If you were asked to intervene by introducing use of insecticide impregnated nets in this community with the aim of showing a 30% reduction of malaria morbidity, knowing that insecticide impregnated bednets have the potential for reducing malaria morbidity by 50% in a community following four years of use, what approach would you follow and what would be your sample size? (consider that money is not a problem)

**Answer:**
Tutor to guide the trainees on the use of Epi-Info or SPSS for calculating sample sizes for the different study approaches.
2.5 Surveys in research

2.5.1 A survey is a cross-sectional study to establish the vital statistics of a population or any other characteristics of the study population at that particular moment. Mention the main characteristics of a survey:

**Answer:**
The main characteristics of a survey or cross-sectional study are that:
- It is conducted at one point in time
- Does not have any follow-up components
- Aims at establishing the current situation.

2.5.2 What are surveys best used for (give four examples)

**Answer:**
Surveys are best used for:
- Establishing vital statistics
- Establishing the size of a particular problem
- Establishing the current characteristics of a population.

2.5.3 What is the relationship between surveys and censuses?

**Answer:**
A survey uses a sample population while a census involves the whole population.

2.6 Surveillance of Malaria

2.6.1 What do you understand by Malaria Surveillance and what does it involve?

**Answer:**
Malaria surveillance includes all efforts and measures which are deliberately taken to acquire knowledge of and monitor:
- The level of malaria transmission
- Changes in the numbers of new cases and deaths due to malaria
- Changes in hospital attendances and admissions
- Efficacy and effectiveness of drugs in use
- any changes in the pattern of the disease in terms of who is affected most.

2.6.2 List different systems of malaria disease surveillance and their use

**Answer:**
These include:
- Epidemic surveillance
- Monitoring of drug resistance levels
- Monitoring hospital attendances, admissions and deaths
- Monitoring mosquito types and densities
- Monitoring infectivity of mosquitoes
- Monitoring correctness of diagnostic measures (quality)
- etc.
2.6.3 What are the most important elements that make disease surveillance useful?

**Answer:**
Important elements of disease surveillance include:

- quality of the diagnostic tools
- quality of information collected
- continuity of the system
- prompt identification of action points and use of such points to trigger action.
Pharmacovigilance and adverse drug reactions (ADRs) monitoring

**Learning Objectives:**

By the end of this unit you will be able to:

- Understand why pharmacovigilance is needed
- Understand the importance of monitoring adverse drug reactions
- Understand the various methods used for ADRs monitoring
- Understand the different type of ADRs and their method of study
- Understand the ADRs reporting system in Tanzania

### 3.1 Pharmacovigilance

3.1.1 What do you understand by Pharmacovigilance?

**Answer:**

This is the science of collecting, monitoring, researching and evaluating data on the effect of medicinal drugs, biological products, herbals and traditional medicines with a view to identifying new information about Adverse Drug Reactions (ADRs) or any other drug related problems and preventing harm to patient.

The framework of pharmacovigilance is broader than that of post marketing surveillance and includes clinical and even pre-clinical development of drugs. The tools used can be epidemiological, experimental (e.g. the attempt to reproduce in animal a type of ADRs in order to understand the mechanism) or diagnostic.

Pharmacovigilance aim at preventing patients from being affected unnecessarily by negative consequences of pharmacotherapy by providing early warning of adverse effects due to drugs.

3.1.2 Give explanation why pharmacovigilance is needed and state its ultimate goal.

**Answer:**

As a rule the available information on medicinal drug collected during the developmental phase prior to marketing is inevitably incomplete with regard to possible adverse reactions and other drug related problems. Often new information comes available during the years after introduction. This is due to the facts that: -
Pre-marketing animal studies have limitation in their abilities to predict human toxicity. Safety data is inevitably incomplete and animal tests are insufficient in predicting human safety.

Pre-marketing human studies (clinical trials) patients are selected and limited in number, the conditions of use in special groups (such as children, the elderly or pregnant women) or drug interactions is often incomplete or not available.

The major changes in the size and nature of the exposed patient population that occur once a drug is available for widespread use emphasise the great importance of adverse event detection and researching.

The ultimate goal of pharmacovigilance is to assist decision-making (e.g. Choice of therapeutic scheme, maintenance or not, of a product on the market).

3.1.3 In your opinion do you think Pharmacovigilance is needed in every country?

Yes [ ] No [ ]

Explain

- YES, because there are differences between countries, and within country (regions) in the occurrence of ADRs and other drug-related problems. This may be because of differences in:
  - drug production
  - distribution and use (e.g. indications, dose availability)
  - genetics, diet, traditions of the people
  - pharmaceutical quality and composition (excipients) of locally produced pharmaceutical products
  - the use of non-orthodox drugs (e.g. herbal remedies which may pose special toxicological problems, when used alone or in combination with other drugs)

3.1.4 What are the objectives and importance of Pharmacovigilance?

Answer:
The objectives of Pharmacovigilance are to
- Ensure long-term safety monitoring in clinical practice and oversee the conduct of clinical trials.
- Identify previously unrecognized hazards and evaluate changes in risks and benefits
- Take action to promote safer use of medicine
- Provide optimal information to patient, clinician and researchers
- Monitor the impact of action taken
- Ensure integrity of research design and protect the trial subjects from potential harm from new drugs.
Importance of Pharmacovigilance:

Is to prevent drug-induced human suffering and to avoid financial risks associated with unexpected adverse effects with the best acceptable safety on use. That is medicines on the market need continuous monitoring in every country.

3.1.5 A new antimalarial drug has recently been manufactured in a country A in Europe. Its safety margin in healthy adults and under five children is excellent, that is there were minimal side effects and no systemic toxicity. The side effects shown were mild headache in 10% of cases in both adults and children and mild diarrhoea in children 5%. Its efficacy in clearing malaria parasites in non immune patients returning from malaria endemic countries is 100%. Would you advice your minister to approve the registration and use of the drug in your country?

Yes [   ] No [   ]

Explain

Answer “no”

The safety data or other information obtained in the country of origin of the drug may not be relevant to other parts of the world, where circumstances may be different. The population characteristics may differ and while there is high probability that the drug may be safe to the new population once it is shown to be safe in the population where it is manufactured, there is no assurance that this will be the case. Testing the safety of new drugs and monitoring their effects at the point of use is useful and may help the early detection of problems related to the drug thus alerting the drug regulatory authorities, physicians, pharmacists, patients, pharmaceutical companies and researchers.

3.1.6 Is Pharmacovigilance a tool for drug quality assurance?

Yes [   ] No [   ]

Answer:

Yes. Pharmacovigilance may help detect the circulation of counterfeit drugs or a batch with lower quality drug. It is an important tool for quality assurance.

3.1.7 Describe the process involved in Pharmacovigilance

Answer:

- Identification: the infrequent reactions, long term reactions, incidence and prevalence of the adverse reactions
- Confirmation: detection of medication error and therapeutic failures as outcomes of ADRs monitoring
- Quantification: evaluation of benefit risk cost ratio
- Implications: the utilisation pattern, overuse, misuse, and study of new indications. Benefit of drugs; especially in long term for prevention of relapse or arresting the progress of the disease
- Preventive measures: - management and communication

27
3.1.8 What are the roles of pharmacovigilance in improving clinical practice?

**Answer:**
- Providing the practitioner with information on the occurring of side effects related to a given prescribed drug
- Provide information on the occurring and extent of Drug Induced Illness
- Improve medical practice
- Inform and determine the need for differential diagnosis and treatment
- Allow early detection and management of drug induced illness

3.2 Adverse Drug Reactions (ADRs) Monitoring

3.2.1 When is ADRs usually monitored?

**Answer:**
- Before the drug is registered and licensed (clinical trials phase I-III)
- After licensing (Phase IV, post marketing surveillance)

3.2.2 What are the differences between ADRs and Side effects?

**Answer:**
An ADRs is a response to a medicine, which is noxious and unintended, and which occurs at doses normally used in man for the prophylaxis, diagnosis or therapy of a disease or for the modification of physiological functions. In this description it is of importance that it concerns the response of a patient, in which individual factors may play an important role. It is a reaction of the body to the drug.

While a side effect is any unintended effect of a drug occurring at doses normally used in man which is related to the pharmacological properties of the drug. Essential elements in this definition are the pharmacological nature of the effect, that the phenomenon is unintended, and that there is no overt overdose.

3.2.3 In monitoring Adverse Drug Reactions (ADRs), what chain of elements is needed to prove evidence of (ADRs)?

**Answer:**
- high level of suspicion of the Adverse Dug Reactions following patients and or practitioners reports
- Evidence from case reports of patients treated with the drug in question
- Evidence from follow up cases treated with the drug and who are not on other medications
- Evidence of allergy or hyper reaction to the drug by a particular group of patients (genetically determined)
- Evidence during drug trial implementation for drugs still on trial.

It is important to remember that for any given ADRs case, there is no certainly that the suspected drug caused the ADRs
3.2.4 List the objectives of ADRs monitoring? Rank them in order of priority

**Answer:**
The direct objectives of monitoring ADR's are:
- To detect the adverse drug reactions as early as possible, especially serious, unknown and infrequently reactions.
- To establish the frequency and incidence of the adverse reactions both the well recognised or newly discovered reactions.
- To identify all factors that may induce and / or influence the development adverse drug reactions (i.e. racial factor, drug interaction irrational drug usage) or affect the severity or incidence.
- To analyse and disseminate information needed in drug prescribing and regulations

3.2.5 Once there is sufficient evidence for Adverse Drug Reactions it is important that the following actions are taken in order to protect potential consumers:
- To provide feedback information on the observed interaction to health professionals.
- To take any regulatory action which may be appropriate
- To issue direct warning to the public, if and when appropriate
- To make essential data available to the analogous system in other counties (via the WHO) so as to promotes the growth of knowledge in this field worldwide.

3.2.6 What is the basic information that need to be born in mind and which needs to be communicated to medical practitioners regarding postmarketing surveillance?

**Answer:**
- That Most drugs are non-toxic, but serious and even life threatening reaction can occur.
- A few drugs have a small margin between the effective and toxic dose.
- All drugs have the potential for causing severe or fatal reactions, even if taken in appropriate doses.
- Some Adverse Drug Reactions may neither be predicted nor avoided.
- Most of the adverse drug reactions are rarely or unlikely to be identified in pre marketing clinical experience of new drugs due to limitations in the sample population
- Researchers and clinicians must therefore be vigilant of Adverse Drug Reactions and unexpected effects.

3.3 The various methods deployed in ADRs monitoring are provided below with a short explanation on each.

3.3.1 Spontaneous Report (SR)
A spontaneous case report represents a diagnosis. It is likely that action will have been taken with that patient on the basis of that diagnosis. Since this diagnosis is made by a highly trained person, there is high probability that it is correct and may represent a case of ADR. However it may grossly under-estimate the actual number of events and could also be wrong. Therefore, spontaneous reporting is not the perfect answer to post-marketing drug surveillance, even though it is cheap, comprehensive over all drugs and uses. This method, has the added advantage of being continuous, and still is, the main way of detecting early or novel ADR signals at the community level.
3.3.2 Prescription Event Monitoring (PEM) and Intensive Medicines Monitoring Programme (IMMP)

Prescription Event Monitoring (PEM) and (Intensive Medicines Monitoring Programme (IMMP), is a way of recording all patients exposed to selected drugs. The patients or their doctors can then be approached by means of a questionnaire to record any or selected events. This approach is particularly useful for new drugs and has the advantage of assembling large cohorts over time. It also allows the follow-up of exposed patient over a long period. This approach provides the possibility of selecting controls from a cohort of users of other drugs allowing the determination of comparative merits of drugs for the same indication. Another strength is that, the method may detect unexpected benefits of therapy.

3.3.3 Post-marketing studies

Special Post-marketing studies on drugs or diseases may yield unexpected signals. On the other hand, such studies are often no larger than pre-marketing studies thus, have little power to find less common ADRs signals. Moreover, they are usually narrowly focused, involve only a few drugs and are limited in duration.

3.3.4 Data linkage

Health-related databases may be very large and contain information on prescribed drugs, indications, concomitant diseases and outcomes of therapy. Since the data are collected systematically, either data linkages or database linkages could provide new ADRs signals. For this to be used for routine signal generation, there would need to be software that would pick out signals according to some predetermined conditions.

3.3.5 Pharmacological studies

Occasionally, animal studies and mechanistic pharmacological studies in humans, suggest the possibility of new ADRs, but these are unusual as primary signals in post-marketing drug safety.

3.3.6 Disease Monitoring

A final possibility for signal generation is to monitor diseases that are often caused by drugs or that are important public health problems. This method has the advantage of continuously monitoring the most important known types of drug morbidity such as agranulocytosis, aplastic anaemia and Steven’s Johnson syndrome. Methods involve running a continuous case-control network for the relevant diseases, using continuously enrolled community controls. The Method is expensive, but the approach aims at detecting the bulk of drug-related morbidity.

3.3.7 Observational Studies

New signals have been found from observational studies. However the signal-generating capacity of observational studies is, both limited and costly. Observational studies play a more important function in the analysis of signals.
3.4 **Adverse drug reactions reporting**

3.4.1 What do you think should be report on ADRs?

**Answer:**
All suspected adverse reactions either known or unknowns, serious or not including minor ones should be reported.
Reports on the new drugs are of great interest because it makes it easier to monitor the performance of these drugs in the country for any suspected ADRs. When an increasing frequency of a given reaction is suspected this is also a reason for reporting.

3.4.2 Who are the appropriate people to report?

**Answer:**
Health care professionals including doctors, dentists, pharmacists, researchers, nurses and other health workers may send ADRs reports to the centre. The zonal Drug Information centres collect ADRs information and send them to the National Centre on weekly basis.

3.4.3 What do you understand by a case report and a signal in relation to ADRs monitoring?

**Answer:**
Case report is a notification relating to a patient with a medical event or laboratory test abnormality describing the history of a patient with a disorder suspected to be drug induced.

A signal is a reported information on a possible causal relation between an adverse event and a drug, the relation being previously unknown or incompletely documented. Usually more than a single case report is required to generate a signal, depending on the seriousness of the event and quality of information.

When different doctors or researchers independently report the same unknown and unexpected/suspected adverse drug reaction, this may be an early and important signal.

3.4.4 When you come across an ADRs what do you think may be an important question you may need to ask yourself. Mention three key questions

**Answer:**
- Can the drug cause the adverse reaction
- Has the drug caused the adverse reaction
- Will the drug cause the adverse reaction
3.5 **ADRs data assessment**

3.5.1 Mention two levels of data assessment in pharmacovigilance

**Answer:**
Individual case assessment,
Aggregated assessment and interpretation

3.5.2 What is it examined and action taken at each level?

**Answer:**
Individual case assessment
- Relevance of observation: unknown, serious, new drug, science, education
- Quality of documentation: completeness, verification, follow-up
- Coding: preferred terms, ACT
- Case casualty assessment

Aggregated assessment and interpretation
- Signal detection
- Regulatory measures
- Serial studies
- Publications

3.5.3 What do you understand by causality assessment?

**Answer:**
Casuality assessment is the process of associating a suspected drug being causing the reaction observed in a patient.

3.5.4 In performing a causalty assessment what are the important criteria would you use in associating the two? Mention four assessment criteria.

**Answer**
- The association in time (place) between drug administration and event
- Pharmacology (features, previous knowledge of side effects)
- Medical plausibility (characteristic, pathological findings)
- Likelihood or exclusion of other causes
Learning Unit 4

Operational research

Learning Objectives

By the end of this unit you will be able to:

- Design an operational research for malarial prevention and control
- Determine the important malaria research questions in health systems
- Determine the important malaria operational research questions in vector control and environmental management
- Determine important operational malaria research questions in Case management.
- Determine the important operational research questions in at risk groups including pregnant women, children below 5 years, refugees or displaced population etc.
- Understand the important anthropological (social science) malaria research questions

4.1 Health System

Consider the Health System in your country.

4.1.1. What are the major problems in the Health System hindering malaria control efforts? List by priority.

The answer may include:

- Frequent drug shortages
- High Costs for utilization of health services
- Limited skills for effective malaria case management at peripheral level (or even central levels depending on the situation)
- Uneven distribution of Health Services between urban and rural populations
- Poor referral systems.

4.1.2 For each priority, what operational research could you consider to solve the problem? Fill in the Table.

Answer:
This will depend on the problems raised and the tutor should guide the individuals/groups on how to formulate research questions (make research titles which have impact on the recipient). Example in the table (see table).
Operational research for malaria control

Problem in Health System

1. Frequent drug shortages
2. Limited skills for effective malaria case management
3. Case management
4. Uneven distribution of Health Services between Urban and rural Communities
5. Poor referral system study to determine the effective of two referral systems in rural communities

Operational Research

- Determining the causes of frequent drug shortages in region x (or country x) and formulate effective solutions.
- Determining the effectiveness of a tailored training programme for malaria case management: A comparative study between intervention & Non intervention.
- Determining the optimum incentives for maintaining functional health services in the rural areas.

4.1.3 Do all communities have easy access to health services in your country? YES/NO

Answer:
No, not all communities have easy access to hospital services.

If NO, How do you assess the access to health services at the community?

Use the table below to fill your answers

<table>
<thead>
<tr>
<th>Communities</th>
<th>Access to Health Services</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excellent</td>
</tr>
</tbody>
</table>

| 1. | 2. | 3. | 4. |

4.1.4 List the operational research to solve problems of access at the community.

1
2
3
4

4.1.5 Discuss the research topics above in terms of:

- Clarity of the issue
- Feasibility of the study
- Applicability of the expected results for solving the problem
4.2 Case Management

4.2.1 What are the important factors which enhance best care and case management?

**Answer:**
- Prompt and effective treatment (effectiveness and quality)
- Low level of drug resistance
- Updated drug policy (effective, accepted and adhered)
- High treatment coverage.

4.2.2 What do you consider as important impediments to case management?

**Answer:**
- Lack of competence of health workers
- Low drug effectiveness (possible high drug resistance, low quality)
- Low compliance (side effects, taste, poor knowledge)
- Outdated drug policy/guidelines/protocols
- Low access to health care/referral services

4.2.3 Therefore in your opinion what should be done to maintain the highest standards in case management?

**Answer:**
- Improve competence of health workers.
- Introduce affordable and effective drugs
- Improve knowledge of effective malaria treatment among communities
- Introduce effective policies
- Increase access to health care and referral services
- Monitor closely the development and spread of drug resistance.

4.3 Vector Control and Environmental Management

4.3.1 List the vector control activities/operations currently being implemented in your country.

**Answer:**
Vector Control activities may include:
- Insecticide Treated Nets (reduce vector human contact)
- Lavicide if done.
- Insecticide indoor spraying if done
- Any environmental management activities etc.
4.3.2 Which among these do you think is most effective and why?

This will depend on the individual answer. In the above case the answer is Insecticide Treated Nets because they are safer than insecticide spraying and highly effective in preventing malaria deaths.

4.3.3 Which one do you think is the least effective and why?

Answer:
In the above case the least effective in the African context is environmental management because of its difficulty implementation and the vast breeding areas in Africa.

4.4 At Risk Groups

4.4.1 What are the most vulnerable groups for severe malaria in your country?

Answer:
- Under five children
- Pregnant women
- Non immune – immigrants from non endemic areas to endemic areas.

4.4.2 What interventions do you think could best help protect these groups? Provide reasons for your reply.

Answer:
- Use insecticide impregnated bed nets
- Prompt diagnosis and treatment
- Intermittent preventive treatment
- Chemoprophylaxis (for immigrants)

The discussion on reasons to be facilitated by the trainer.

4.4.3 Discuss the epidemiology of malaria in infants and list a number of interventions that you think will be effective.

Answer:
Infants in malaria endemic areas are born with partial immunity to malaria. This allows them to be infected without developing malaria symptoms and it also gives them the ability to clear parasites.

However this immunity only lasts for three months after which they become most vulnerable to malaria attacks. The consequence of malaria at this age may be severe anaemia, convulsions at death.

After one year of age, they start developing own partial immunity which at the age of five years is strong enough to prevent frequent bouts of malaria and its severe consequences.
Interventions which may be effective to prevent severe consequences of malaria in this age groups include

- use of insecticide impregnated bed-nets
- intermittent preventive treatment
- introduction of a vaccine at the age of three months (were a vaccine available)
- ensure prompt treatment whenever a child falls ill.

4.4.4 For each Intervention listed above, indicate on a line scale when it would be most appropriate to introduce the intervention. Explain why?

**Answer:**

<table>
<thead>
<tr>
<th>Bednet</th>
<th>Vaccine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prompt Rx</td>
<td>ITNs</td>
</tr>
</tbody>
</table>

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |

- Bednet and prompt treatment right from birth.
- Vaccine and intermittent preventive treatment at 3 months of age.

4.4.5 Discuss the epidemiology of malaria in pregnancy and list a number of interventions which you think would be effective in reducing the frequency and consequences of severe malaria in pregnancy.

**Answer:**

- Malaria in pregnancy is severe especially in primigravidae.
- Placenta parasitization result in low birth weight or may cause abortions.
- Pregnant women may become severely anaemic and die.
- Low-birth weight is a risk for neonatal death.

Interventions in Pregnancy may include:

- Use of Insecticide Treated Nets.
- Use of Intermittent Preventive treatment.
- Prompt and correct treatment whenever ill.
- Vaccine when available (Hope for the future)
4.5 Malaria Diagnosis General

4.5.1 Is malaria diagnosis a problem in your country? YES/NO
If YES, explain why?
The answer to 6.4.1 will depend on the country situation. In most cases it is a problem because there are no easy cheap and specific diagnostic tools at community level.

4.5.2 List the major problems in malaria diagnosis and the targeted operational research by filling the table below.

<table>
<thead>
<tr>
<th>Diagnosis Problem</th>
<th>Operational Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lack of diagnostic services at peripheral level</td>
<td></td>
</tr>
<tr>
<td>2. Lack of quick, cheap and specific diagnostic tools for are at community level</td>
<td></td>
</tr>
<tr>
<td>3. Lack of diagnostic skills to health workers working in rural health services/or even urban</td>
<td></td>
</tr>
<tr>
<td>4. Lack of screening tools and programmes</td>
<td></td>
</tr>
<tr>
<td>5. Difficult access to health services for diagnostic purposes.</td>
<td></td>
</tr>
</tbody>
</table>

4.6 Community Access to drugs

4.6.1 Do communities have the same level of opportunity to have access to antimalarial drugs in your country? YES/NO

Answer:
Some communities have more difficulties of access to drugs.

4.6.2 If NO, could you explain with the aid of a table or graphical presentation the different level of access to anti-malaria drugs?

<table>
<thead>
<tr>
<th>Type of Community by Levels</th>
<th>Accessibility to drug</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excellent</td>
</tr>
<tr>
<td>Urban Communities</td>
<td></td>
</tr>
<tr>
<td>Rural Communities</td>
<td></td>
</tr>
<tr>
<td>Coastal Regions</td>
<td></td>
</tr>
</tbody>
</table>
4.6.3 What are the major problems and indicate the appropriate operational research to solve the problem.

<table>
<thead>
<tr>
<th>Drug Access Problem</th>
<th>Operational Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
</tr>
</tbody>
</table>

4.7 Anthropological Social Cultural Belief and Altitudes

4.7.1 List the Cultural beliefs which impede malaria control and targeted operational research.

**Answer:**
Cultural beliefs may include:
- Belief that anaemia and convulsions are not malaria related but rather spiritual problems.
- Belief that hospital treatment for malaria in pregnancy is harmful.
- Belief that malaria is caused by witchcraft etc.

4.7.2 List the community attitudes that impede malaria control efforts and operational research.

**Answer:**
- Fear of finger prick diagnosis and hence not sending sick children to hospital (health services).
- Delays in referring sick children for treatment even when very sick for fear of modern medicine.
- Acceptance that malaria kills and there is nothing one can do about it etc.

4.7.3 What are in your opinion the major socio-cultural impediments towards acceptance and use of malaria control measures?

The **answers** should target
- Taboos
- Beliefs
- Practices
- etc
4.7.4 What would facilitate access and use of malaria control measures? Give examples in your explanation.
For group discussion

4.8 Malaria epidemics

4.8.1 What is a malaria epidemic?

Answer:
A sharp increase in malaria incidence among populations in whom the disease is infrequent, or an increase in clinical malaria in areas of moderate transmission constitutes an epidemic.

Malaria epidemics occur principally in areas of low transmission, where no single age group in the population is immune. The introduction of malaria, particularly if exacerbated by changes in rainfall and temperature, can trigger explosive epidemics that affect both adults and children. However, epidemics can also occur in areas of higher transmission as a result of the abandonment of control programmes and sudden increase in transmission and arrival of non immune immigrants.

4.8.2 What are the conditions which favour malaria epidemics

Answer:
- Increase in vector density and infectivity
- Increase in transmission
- Climatical condition providing significantly more breeding areas
- Displacement of populations from non endemic to endemic areas.
- New settlements – solders or farmers

4.8.3 Who are the most vulnerable groups to severe malaria in an epidemic situation?

Answer:
- Young Children - In high endemic areas
- Pregnant women - In high endemic areas
- Everybody if non immune

4.8.4 What are the important preventive and control measures for malaria epidemics?

Answer:
- Identification of high risk areas
- Monitoring of malaria cases continuously
- Integrated disease surveillance that takes good consideration of malaria
Learning Objectives:

By the end of this unit you will be able to:

- To understand the types and application of data collection methods
- Define different types of variables
- Understand how to manage data appropriately (processing, storage and retrieve)

5.1 Data collection

5.1.1 What is data? Explain the main differences between nominal, ordinal and numerical data.

**Answer:**
Data is a collection of items of information/package of information or facts which is used to draw inferences for decision making and planning.

Nominal date is data, which is arranged in unordered qualitative categories instead of individual absolute values. This includes race, religion, ethnicity etc.

Ordinal data is data arranged or classified in ordered qualitative categories. Example: Social class I, II, III, referring to High, Middle and low class. Example: Parity I, II, III, IV etc.

Numerical data is data, which is presented in numbers or categories of numbers. Example Age 1, 2, 3, 4 etc.
Age 1 – 5 = 1 6 – 10 = 2, 11 – 15 = 3 etc.

5.1.2 Mention the data collection materials or tools you know about?

Data collection methods or tools include questionnaires (structured, semi structured) special forms like specimen or laboratory forms, records, interviews etc.

5.1.3 In what form can research data be recorded and stored?

**Answer:**
Research data is usually recorded and stored in databases.
5.1.3 What is the essential process that raw data should undergo before it is recorded?

**Answer:**
The Essential and stored processes that raw data must undergo before it is recorded include:
- Checking for accuracy and consistency
- Checking for completeness of information

5.2 Variables

5.2.1 What is a variable?

**Answer:**
A variable is any attribute, phenomenon or event that can have different values. It is recorded as a quantity that has variations and determines a characteristic of interest in a study.

5.2.2 Mention the different types of variables?

**Answer:**
Variables can be classified as:
- Qualitative or categorical and quantitative or numerical.
- Qualitative variables are non-numerical like place of birth, ethnic or religious group, type of work etc and can be of a special type called binary variable which has two categories like sex Male or Female.

5.2.3 Explain the relationship between different variables.

**Answer:**
Qualitative variables can be either discrete or continuous.
- Discrete variables have only whole number measurements eg. Monthly or daily malaria cases, deaths, births etc.
- Continuous variables are measured in continuous numbers, that is with decimal points like weight, age, volume, parasite density, haemoglobin level etc.

Give an example of such relationship

**Answer:**
The relationship between variables is explained in the way they relate to particular outcome or event. In this sense, variables can be classified into dependent and independent variables.

Independent variables are those which in a given relationship, do not depend on the other variable to exist.
5.2.4 What is a dependent variable?

**Answer:**
Dependent variables are those which in a particular relationship, depend entirely on the status or position of the other variable to have a certain value.

5.2.5 What is an independent variable?

**Example:**
In the relationship between Age and parasite density, Age is an independent variable while parasite density is a dependent variable. This is because parasite density depends on the level of acquired immunity which is age related but age of an individual has nothing to do with parasite density for any of its values.

5.3 Data Processing

5.3.1 Which data bases are you aware of? Which of those are you able to operate?

The **answer** may be SPSS, EPI-INFO, STAT, FOXPRO, Excel spreadsheet etc. Trainees will indicate which ones they are familiar with and trainers have to take note in order to target support accordingly.

5.3.2 What is essential to ensure data quality in the data entry process and why?

**Answer:**
Data quality is ensured in the process of data entry by instituting a system of independent double entry. This is a system in which each data is entered by two different individuals independently. It bears its strength in the assumption that two independent individuals are unlikely to make the same mistake in interpreting or typing the keys of a computer. Comparison of the two entries therefore reveal the mistakes made by the different individuals during the data entry process.

5.3.3 What is data cleaning/editing and why is it important?

**Answer:**
Data cleaning/editing is the process of making checks on the data by comparing the double entries, looking for outliers or unrealistic values and multiple identities for some individual etc.

It is a process which ensures data quality before analysis.
Learning Objectives

By the end of this unit you will be able to:

- To understand the different frequencies in qualitative data and their application
- To understand the different frequencies in quantitative data and their application
- To understand the different measures of central tendency and their application

Data analysis involves the arrangement of data in such a way that it explains a certain feature of the issue being studied.

It should be done in the most appropriate way and as simply as possible.

What is important is to provide a clear and reliable answer to the research question.

Data is usually presented in variables. The values or counts of a particular variable are called frequencies. Frequencies indicate how often a particular value of a variable occurs in a given data. Each individual count is a frequency. Frequencies can be derived from both quantitative and qualitative data.

6.1 Frequencies in qualitative data:

The table below shows the level of knowledge about the correct treatment for malaria among cotton factory workers in a hypothetical town in Tanzania, living under intense and perennial malaria transmission.

Table 6.1 Relative frequencies of the level of knowledge of the correct treatment for malaria among cotton factory workers

<table>
<thead>
<tr>
<th>Cotton Factory Workers</th>
<th>N</th>
<th>Level of Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Males</td>
<td>126</td>
<td>28</td>
</tr>
<tr>
<td>Females</td>
<td>124</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>350</td>
<td>32</td>
</tr>
</tbody>
</table>

Using the data in table 6.1, calculate the relative frequencies of knowledge of the cause of malaria in the cotton factory workers.
6.1.1 Frequencies and relative frequencies can be presented in the form of illustrations called bar charts, pie charts and histograms.

Use the data in Table 8.2 below to present the relative frequencies in a bar chart, and pie chart. Which one of the illustrations presents this data best? Could this data be presented in a histogram?

**Table 8.2 Distribution of the outcome of 100 malaria cases treated in dispensaries with chloroquine, by village of residence**

<table>
<thead>
<tr>
<th>Residential Village</th>
<th>Cured</th>
<th>Not cured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ilala</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>Buguruni</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Chang’ombe</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>Upanga</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Magomeni</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 8: Proportions of the outcome of 100 malaria cases treated in Dispensaries with chloroquine, by village of residence**

Note: Important considerations in forming frequency distribution are whether the data is best presented as a frequency distribution or relative frequency and to identify the dependent and the independent variable. In a frequency distribution, the dependent variable is presented in the Y axis while the independent variable forms the X axis.
Figure: Proportions of the outcome (cured by village of residence and overall not cured) of 100 malaria cases treated in dispensaries with chloroquine

Figure: Proportions of the outcome (cured) of 100 malaria cases treated in dispensaries with chloroquine by village of residence
6.2 Frequencies in quantitative data

The following table shows the hypothetical number of malaria patients aged 0-36 months who were seen at the outpatient clinic of Idete Hospital and those who were admitted to the same Hospital in the period January to June 1999.

Table 7.2

<table>
<thead>
<tr>
<th>Age</th>
<th>Out Patient Cases</th>
<th>In Patient Cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-6 months</td>
<td>42</td>
<td>32</td>
<td>74</td>
</tr>
<tr>
<td>7-12 months</td>
<td>176</td>
<td>83</td>
<td>259</td>
</tr>
<tr>
<td>13-18 months</td>
<td>170</td>
<td>50</td>
<td>220</td>
</tr>
<tr>
<td>19-24 months</td>
<td>168</td>
<td>32</td>
<td>200</td>
</tr>
<tr>
<td>25-30 months</td>
<td>130</td>
<td>25</td>
<td>155</td>
</tr>
<tr>
<td>31-36 months</td>
<td>45</td>
<td>8</td>
<td>53</td>
</tr>
</tbody>
</table>

6.2.1 What type of data is presented in table 8.2? Indicate the independent and dependent variables in the data.

The data in table 8.2 is formed of continuous variables. Any number of children may attend on a day (within the age groups)

Age is the independent variable

Out patients and In patients are the dependent variables

6.2.2 What is a histogram and how does it differ from a bar chart? Draw a histogram representing the data.

Answer:

Histogram is a frequency distribution of continuous data presented bars. The different between a bar chart and a Histogram is that bar charts are used for discrete variables while Histograms are used for continuous variables.

In a Bar chart the bars are separate and independent of each other. In a Histogram they are joined together
6.2.3 Discuss and explain what is a frequency polygon

**Answer:**

A frequency polygon is a frequency distribution illustrated by a continuous line joining the midpoints of the categories (Histograms).
Figure 4: Frequency polygon of outpatient and inpatient attendances at Idete Hospital by age groups

Example:

IPC: In-patient Cases  OPC: Out patient cases

6.2.4 As in figure 4, frequency polygons are best used to compare frequency distributions between populations. 
Explain the different types of frequency polygons by shapes

a) Symmetrical distributions are frequency polygons which have symmetrical shapes. That is can be divided into two equal parts by a line passing through a middle point showing that the lower and higher variables are equally distributed.

Example:
b) Skewed distributions are frequency polygons in which the variables are not evenly distributed, and in which the greater number of variable (the peak) is either shifted to the left or to the right. The distribution with therefore have a tail on either the lower or the upper side

When the tail is in the left or lower side it is described as skewed to the right. When the tail is in the right side it is described as skewed to the left.

Examples:

Data skewed to the right

![Data skewed to the right](image)

Date skewed to the left

![Date skewed to the left](image)

c) Bimodal distribution presents two peaks in the distribution which may be at the same or different levels.

![Bimodal distribution](image)
d) Reverse J shaped distributions express the relationship where the values of the variable are present in a reverse J shape. This is characteristic of life tables or survival curves.

\textbf{Reverse J Shaped frequency polygon}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{reverse_j_shaped.png}
\caption{Reverse J Shaped frequency polygon}
\end{figure}

\vspace{1cm}

e) Uniform distribution is when the dependent variables present the same value over different ranges of the independent variable.

\textbf{Example:}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{uniform_distribution.png}
\caption{Uniform distribution}
\end{figure}
6.3. Measures of Central Tendency

6.3.1 What do you understand by measures of central tendency?

Answer:
Measures of central tendency indicate how the variables are distributed in relation to the midpoint value or centre. In a normal distribution, the reference point is the mean.

6.3.2 Discuss and explain what the Mean is?

Answer:
The mean or average is the ratio of the sum of all the values of a variables divided by the number of variables. This may also be called the arithmetic mean.

Another kind of mean is the geometrical mean which is preferred when the distribution is positively skewed or skewed to the right.

Symbols:
The Arithmetic mean = \( \frac{\sum x}{N} \)

Geometric Mean = \( n \sqrt[n]{(x_1)(x_2)x_3 \cdots x_n} \)

Such that log GM = \( \frac{\sum \log x}{N} \)

6.3.3 Describe and explain what is Median?

Answer:
The Median is the central value. In a symmetric distribution, this is equal to the mean. In a skewed distribution this is away from the mean either to the right or left depending on the skewness.

6.3.4 Describe and explain what is Mode?

Answer:
The mode is the most frequent value in a frequency distribution.

6.3.5 The following data was collected from 12 children in each age group for age one to five years.

Children aged 1 year presented the following densities per microlitre: 25360; 18300; 17289; 16594; 15648; 15648; 14594; 10494; 11370; 9300; 8596; 7295.

Age 2 years: 3671; 3052; 2817; 1948; 4267; 2481; 2500; 6372; 9520; 7415; 5439.

Age 3 years: 3274; 4152; 2350; 1950; 1500; 1648; 1500; 1970; 2149; 1500; 3758, 3558.

Age 4 years: 996; 1105; 1376; 1011; 1085; 926; 865; 895; 895; 3650; 1520; 1484.

Age 5 years: 744; 705; 655; 682; 612; 534; 487; 568; 605; 568; 545; 430;
Calculate the mean parasite density per age?

The mean parasite density is
\[ \frac{\sum x}{n}. \]

- For Age 1 = 14207.33
- For Age 2 = 4410.67
- For Age 3 = 2442.42
- For Age 4 = 1317.33
- For Age 5 = 594.54

6.3.6 What pattern do you get? (Draw a graphical presentation to help you)

**Mean parasite density decreases with age**

![Graph showing mean parasite density decreases with age](image)

6.3.7 Could you provide a possible explanation for the observed pattern with the knowledge that the data is from an area of high and perennial malaria transmission.

**Answer:**
People living in malaria endemic areas develop partial immunity which increases with the time of exposure and hence with age.

The partial immunity does not present infection but suppresses the level of parasite density so that the older the person is the lower level of parasite densities they develop.

6.3.8 Identify the median and mode values of parasite density for each age groups

<table>
<thead>
<tr>
<th>Age</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 1</td>
<td></td>
<td>15648</td>
</tr>
<tr>
<td>Age 2</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Age 3</td>
<td></td>
<td>1500</td>
</tr>
<tr>
<td>Age 4</td>
<td></td>
<td>895</td>
</tr>
<tr>
<td>Age 5</td>
<td></td>
<td>568</td>
</tr>
</tbody>
</table>
6.3.9 What can you say about the median and mode and their relationship to the mean in this case?

**Answer:**
The Median and Mean are not at the same level (do not have same value). Therefore the distributions are not symmetrical.
The values of the mode and the median also decrease with age as the mean does.

6.4 **Measures of variation - Range, Variance, Standard Deviation and the Coefficient of variation**

6.4.1 The Range
- What is the range?
- Calculate the range of parasite density for each age group in the data provided in 8.4 above

Range = Highest value - lowest value

In the example given in 8.4, the ranges of parasite densities in children within the different age groups.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Highest Value</th>
<th>Lowest Value</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25360</td>
<td>7295</td>
<td>18065</td>
</tr>
<tr>
<td>2</td>
<td>13671</td>
<td>4267</td>
<td>9404</td>
</tr>
<tr>
<td>3</td>
<td>4152</td>
<td>1500</td>
<td>2652</td>
</tr>
<tr>
<td>4</td>
<td>3650</td>
<td>865</td>
<td>2785</td>
</tr>
<tr>
<td>5</td>
<td>744</td>
<td>430</td>
<td>314</td>
</tr>
</tbody>
</table>

6.4.2 The variance:
Define the variance and calculate the variance of the parasite density in the age group 1 year in 8.4

**Answer:**
The variance is a measure of the deviations of the observations from the mean. It is
Sum of the squares of the differences between the value $X$ and the mean $\overline{X}$ divided by the degree of freedom

$$Variance = \frac{\sum (x - \overline{x})^2}{(n - 1)}$$

The degrees of freedom are calculated from the number of observations $n$, minus one ($n - 1$).

The variances of the parasite densities within the age group 1 year in 8.4 can be calculated as:
6.4.3 The Standard Deviation (SD) and the Coefficient of variation b (CV).

The standard deviation (SD) is a measurement, which describes how observations are positioned around the mean. It is the most frequently used measure of spread in the medical field.
It is used in the assumption that the variables (observations) present a normal distribution.

Discuss the Standard Deviation and its relationship to the variance.
Calculate the standard deviation of the parasite densities in the group aged 1 year in 8.4 above.

The standard deviation is simpler to understand than the Variance because it converts the deviations into the original measures. It can be defined as the squire root of the Variance.

\[
SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}
\]

The standard deviation of the parasite densities of the group aged 1 year can therefore be calculated as follows:
Take the variance calculated in 8.5.1 above = 285813932/11

Get the square root of the variance = \(\sqrt{2598308472}\)

\[
SD = 5097.36
\]
6.4.4 How does the Standard Deviation relate to the Normal Distribution?

**Answer:**
The area under the normal distribution is measurable in relation to points away from the mean. These points are multiples of the standard deviation indicating the position of a variable around the mean in a normal standard distribution.

6.4.5 Discuss and describe how the sample mean relates to the actual population mean.

**Answer:**
The importance of a sample is how it provides information on the whole population, from which it was derived.

Since we do not know the population mean and its standard deviation, the sample mean and standard deviation are used to estimate the population standard deviation and mean.

Because of variations in sampling procedures, the sample mean is unlikely to be exactly equal to the population mean. However, if we collected many repeated samples of a population, and posted the distribution of the means of these samples, the mean of such a distribution would be very close or equal to the population mean.

In general, we know that there is 95% probability that the population mean lies between $\bar{X} \pm 2SD$.

From the example in 8.5.2 above we can derive that the population mean for the age group 0-12 months lies between 4012.61 and 24402.05.
Learning Objectives:

By the end of this unit you will be able to:

- Describe the Normal distribution curve and its difference from other distributions
- Explain the position of the mean, median and mode in a normal distribution curve
- Understand what are confidence intervals and their application
- Understand how to test the working hypothesis in reference to the Null Hypothesis.
- Use the Z, t, $X^2$ and Sensitivity testing

7.1 The Normal Distribution Curve

7.1.1 What do you understand by the normal Distribution Curve?

**Answer:**

The normal Distribution curve often also called the Gaussian distribution is a smooth and continuous bell-shaped distribution, symmetrical about the mean. It is a probability distribution and therefore the total area under the curve is equal to 1 since, the sum probabilities for a given set of events is equal to one.
7.1.2 Explain the position of the Mean, Median and Mode in a Normal Distribution curve.

7.2 The Standard Normal Distribution, the standard normal deviate and the area under the normal distribution curve.

The standard normal distribution curve, also known as the Z distribution, is not only used to determine the position of a given value under the curve (or Z), but also the area between the position of a +z and a –z value. In this distribution, changes in the mean only pushes the curve more to the right or the left of the axis while changes in the standard deviation only makes the curve wider or narrower, and affects its height. The standard normal distribution has a mean of zero and a standard deviation of 1.

Any normally distributed variable can be related to the standard normal distribution by converting it into the so called “z” value. This is derived in the following steps:

Step No.1 Subtracting the mean from the value of the variable x
\[ x - u \]
Step No.2 Dividing the value obtained in step No.1 by the standard deviations
\[ \frac{x - u}{s} \]

Thus the equation for Z is:

\[ z = \frac{x - u}{s} \]

Where
- x = the variable
- u = the mean
- s = the standard deviation

z is called the Standard Normal Deviate (SND)

The Standard Normal Deviate (SND) expresses the value of the variable in terms of the number of standard deviation it is away from the mean.

When for a given variable the value of z = 1 it means that the variable is one standard deviation away from the mean (on the right hand side of the curve). If the value of z = -1 it means the variable is one standard deviation away from the mean on the left side of the curve.

One of the most important reference points in the Standard Normal Distribution is the 1.96 Standard deviations away from the mean which encompasses 95% of the area of the Standard normal distribution curve between –z and z.

Therefore the 1.96 point of z is called the five percent point of the Normal distribution. At this point 2.5% of the distribution lies further than 1.96 standard deviations on each side of the distribution.

In a standard normal distribution, 95% of the sample means should lie within 1.96 standard errors above and below the population mean.
It follows therefore that the interval from \( x - 1.96 \text{ se} \) and \( x + 1.96 \text{ se} \) represents likely values for the population mean and is called the 95% confidence interval for the population mean whereby \( x - 1.96 \text{ se} \) and \( x + 1.96 \text{ se} \) are the lower and upper 95% confidence limits of the population mean respectively.

The other reference point is the 99% confidence interval which is \( x \pm 2.58 \text{ se} \) from the population mean.

Normally the reference points are applicable in large samples (>60) and when the distribution is normal. When smaller samples are concerned in otherwise a normally distributed populations the \( t \) – distribution is used.

Mathematical tables are available which include the standard normal distribution table and can be used to determine the areas under the normal distribution curve.

7.2.1 Confidence Intervals and Confidence Limits

Many simple analysis or calculations will provide a point estimate. This would be satisfactory if we would be considering the whole population and not samples. However, because we are deriving our estimates from a sample of the population, there is need to indicate the acceptable limits for the variability of our sample mean. This is called the CONFIDENCE INTERVAL.

Example: If our estimated mean parasite density is 5,000 parasites per microlitre, we should be able to state with confidence that the true mean is expected to lay say between 4,000 and 6,000 parasites per microlitre. This interval includes our estimated mean.

In calculating the confidence interval for a sample mean, the main assumptions are that:
- The sample means are normally distributed
- The sample mean is the population mean \( \mu \) and the standard deviation is equal to the standard error of the sample means \( \sigma/\sqrt{n} \)

In order for the above assumption to be true or close to the truth, the sample size must be large such that \( n \) is at least equal to 60. Since we know that 95% probability that the sample mean lies within \( \pm 1.96 \text{ S.E.} \) fro the population mean. Therefore the interval \( \bar{x} \pm 1.96 \) contains the location of the population mean. This interval, which contains 95% probability that the population mean is found within it, is called the 95% confidence interval.

For large samples, this confidence interval is given by the equation:

\[
95\% CI = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}
\]

The most commonly used confidence levels are 90%, 95% and 99%.

Tutor to explain the \( t \) distribution and use of \( t \) distribution table.
To calculate the confidence intervals for other levels rather than the 95% level we substitute 1.96 by the value of 2 in the standard normal distribution.

Example: 2 for 99% confidence level is 2.58.

Therefore: 

\[ 99\% CI = \bar{x} \pm \left( 2.58 \times \frac{\sigma}{\sqrt{n}} \right) \]

\[ = \bar{x} \pm 2.58 \times SE \]

7.3 Hypothesis Testing

In research we usually test the researcher’s hypothesis or the working hypothesis against the null Hypothesis. As seen before, the two Hypotheses are against each other and our test statistics help us to reject the Null Hypothesis or declare failure to reject the Null hypothesis even though this does not mean accepting the Null Hypothesis.

In testing the working hypothesis, there are three important elements which we must consider namely:
- The power
- The errors in the tests
- The P. Value

**Hypothesis testing and possible outcomes**

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accept the null hypothesis</strong></td>
<td>Correct decision</td>
<td>Type II</td>
</tr>
<tr>
<td><strong>Reject the null hypothesis</strong></td>
<td>Type I</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>

7.3.1 The Power

Power is the probability of rejecting the Null Hypothesis when it is indeed false. It is the strength with which we are able to conclude that the working hypothesis is true when it is really true.

It is calculated as (1-B). In other words it is subtracting the type I error from one.

Therefore there are three possibilities to the final conclusions of our test.
Possibility No.1: Is that we reject the Null hypothesis when it is actually true. This is what we intended to do in our working hypothesis and we are actually able to show that the Null Hypothesis is wrong, when it is actually true. Here we commit type I error.

**Figure:**

Reject Null Hypothesis  
If sample mean here  
(A)

Accept Null Hypothesis  
if sample mean here  
(C)

Reject Null Hypothesis  
Sample mean here  
(B)

Possibility No.2: Is that we are not able to reject the Null Hypothesis when it is actually false. This means our test was not robust enough and we are not able to detect the difference which actually exists. This type of error is called type II error or $\chi$ error.

**Figure:**

Accept Null Hypothesis  
If sample mean here  

Reject Null Hypothesis  
If sample mean here  

$\mu$  
$\mu^1$  
$b\%$
Possibility No.3: Is that we accept the Null Hypothesis when it is actually true. In this case there is actually no significant differences between what we are testing and we are able to show that this is the case.

Possibility 4. Is that we reject the null hypothesis when it is actually false.

This is the case when we have enough power in our sample calculation

In sampling we aim to have enough power to prevent making both error I and II affirm that there is a true significant difference in our measures. When we fail to affirm so, we are not sure whether it is because this is true as possibility 4 above or it is because we do not have the power to detect the difference. We therefore just conclude that we could not show statistically significant differences between what we were comparing.

7.3.2 P-Value

P-Value is the probability of obtaining a result as extreme as or more extreme than the one observed.

P-Value indicates the status of the findings or to whether a significant difference has been shown or not.

The reference point is 0.05. Any P-value greater than 0.05 indicates that there is no statistical difference between the compared values.

Any P-value equal to or less than 0.05 indicates that there is a significant difference between the compared values.
7.4 The Z-test for means

The Z-test is applicable when we are dealing with a standard Normal Distribution with a mean \( \mu \) and a standard deviation of 1. Such a curve is called the Z-distribution (Figure below).

\[
\text{Z} = \frac{X - \mu}{\delta}
\]

It must be noted that for a normal distribution:
- The mean ± 1 standard deviation contains 66.7% of the area under the normal curve.
- The mean ± 2 standard deviations contains approximately 95% of the area under the curve and
- Mean ± 3 standard deviation contains 99.7% of the area under the normal curve.

The value of Z is expressed as the difference between the sample mean and the population mean, divided by the standard error

\[
Z = X - \mu
\]

7.5 The t-test

This is a text used for t distribution since we usually do not know the population mean and the population standard deviation.

In using the sample mean we are required to transform the distribution into a t-distribution. This is also symmetrical, has a mean of zero but its standard deviation is greater than 1. It is therefore more broadly distributed around its mean.

T-test is also used when we are dealing with small samples < 30. However when the sample is big enough, the t distribution becomes narrower and approaches that of the normal distribution.

\[
\text{The value of } t = \frac{X - U}{\text{SD/den}} = \frac{X - U}{\text{SE}}
\]

Note: Tutors to show trainees how to use the t distribution table
7.6 The $X^2$ test

This is also known as the test of Goodness of Fit.

The $X^2$ test deals with comparisons between the observed frequency or mean and the expected frequency or mean.

Data to be analyzed by $X^2$ is best understood if it is arranged in a table where the columns show the frequency of observations (both observed and expected) and the rows represent the outcome of interest like parasite cleared fully and parasite not fully cleared.

Table: The frequency of parasite clearance when children were treated by Drug A or B for malaria.

<table>
<thead>
<tr>
<th>Status</th>
<th>Observed Frequencies</th>
<th>Treat A</th>
<th>Treat B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parasites Cleared</td>
<td></td>
<td>40</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Parasites Not Cleared</td>
<td></td>
<td>60</td>
<td>90</td>
<td>150</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
</tbody>
</table>

First we need to calculate the expected frequencies which could occur by chance. This is done by taking for each observed value the column totals multiplied by the row totals and divide the result by the Grand total.

Example: For treatment A.

The expected value of Parasite clearance is \( \frac{100 \times 50}{200} = 25 \)

Similarly for Drug B the expected parasite clearance is \( \frac{100 \times 50}{200} = 25 \)

You will note that the expected value is the same for the different drugs because we are dealing with chance and in our Null Hypothesis we expect no difference in outcome.

Once these values are calculated you get a table like the one below.

<table>
<thead>
<tr>
<th>Status</th>
<th>Expected Frequencies</th>
<th>Drug A</th>
<th>Drug B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parasite Clearance</td>
<td></td>
<td>25</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>No Parasite Clearance</td>
<td></td>
<td>75</td>
<td>75</td>
<td>150</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
</tbody>
</table>
The $X^2$ is obtained by taking the observed value minus the expected and dividing this by the expected.

$$X^2 = \sum \frac{(O - E)^2}{E}$$

When there is no difference in outcome $X^2$ is small. When there is a significant difference $X^2$ is large.

Make the calculations in the table above

The cut off point is 3.841. Values below this indicate chance effect. Values greater than this number indicate statistical significant in difference of outcome. The higher the number the larger the difference.

### 7.7 Sensitivity Testing

#### 7.7.1 What do you understand by sensitivity of a test?

**Answer:**
Sensitivity is the proportion of true positives that are correctly identified as such. In mathematical terms it is 1 minus false negative rate

#### 7.7.2 Explain the specificity of a test

**Answer:**
Specificity is the proportion of true negatives that are correctly identified as such. In mathematical terms it is 1 minus false positive rate

#### 7.7.3 Explain what a false positive is.

**Answer:**
False positive is the proportion which has been identified by the screening test as being positive when actually it is negative by the standard diagnostic test.

#### 7.7.4 Explain what a false negative is.

**Answer:**
False negative is the proportion which has been identified by the screening test as being negative when actually it is positive by the standard diagnostic test.

#### 7.7.5 Explain what is meant by predictive value of a positive test

**Answer:**
It is the proportion of true positives among those who were identified by the screening test as being positive.
7.7.6     Explain what is meant by predictive value of a negative test

**Answer:**
It is the proportion of true negatives among those who were identified by the screening test as being negative.

**Use a table to show**

**Give examples**

<table>
<thead>
<tr>
<th>Screening test</th>
<th>Standard Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Malaria +ve</td>
<td>Malaria –ve</td>
</tr>
<tr>
<td>Malaria +ve</td>
<td>68</td>
<td>112</td>
</tr>
<tr>
<td>Malaria –ve</td>
<td>132</td>
<td>688</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>800</td>
</tr>
</tbody>
</table>

False positives = 112  
False negatives = 132  

**Calculate:**

a) **Sensitivity** = \( \frac{68}{200} \times 100 = 34\% \)  
   = 1 – \( \frac{132}{200} \times 100 \)  
   = 100\% - 66\% = 34\%

b) **Specificity** = \( \frac{688}{800} \times 100 = 86\% \)  
   = 1 – \( \frac{112}{800} \times 100 \)  
   = 100\% - 14\% = 86\%

c) **Positive predictive value** = \( \frac{68}{180} \times 100\% \)  
   = 37.8\%

d) **Negative predictive value** = \( \frac{688}{820} \times 100\% \)  
   = 83.9\%

**7.8 Summary Methods for Multivariate Analysis**

**7.8.1 For two variables**

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Dependent variable</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Nominal</td>
<td>Chi-Square ( (X^2) )</td>
</tr>
<tr>
<td>Nominal (binary)</td>
<td>Numerical</td>
<td>Student t – test</td>
</tr>
<tr>
<td>Nominal (more than 2 values)</td>
<td>Numerical</td>
<td>One-way ANOVA</td>
</tr>
<tr>
<td>Numerical</td>
<td>Numerical</td>
<td>Regression</td>
</tr>
</tbody>
</table>
7.8.2 For multiple variables

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Dependent variable</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Nominal</td>
<td>Log-linear</td>
</tr>
<tr>
<td>Nominal and numerical</td>
<td>Nominal (dichotomous)</td>
<td>Logistic regression</td>
</tr>
<tr>
<td>Nominal</td>
<td>Numerical</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Numerical</td>
<td>Numerical</td>
<td>Multiple regression</td>
</tr>
<tr>
<td>Nominal with confounding factors</td>
<td>Nominal</td>
<td>Mantel-Haenszel</td>
</tr>
</tbody>
</table>
Learning Unit 8

Data Analysis III : Multivariate analysis

Learning Objectives

By the end of this unit you will be able to:

- Understand and apply the significance tests for Multivariate analysis
- Identify appropriate packages for Multivariate analysis
- Identify the type of data needed for Multivariate analysis
- Interpret the results of different Multivariate methods of analysis

8.1 Correlations

These are studies which are used to describe the relationship between an independent or explanatory variable and a dependent or outcome variable

8.1.1 What illustrations are best in showing the correlation between a dependent and independent variable?

Answer:
Scatter diagrams or scatter plots are used to show correlations between independent and dependent variables.

Example:
The relationship between years of continuous stay in a malaria endemic area and the number of malaria episodes per year may hypothetically be presented as in the following scatter diagramme
Correlation coefficient $r$ measures the precision of the linear relationship between an independent and a dependent variable. It is the slope of the line best joining the points of the scatter diagram or plot of the values of the variables.

The value of $r$ ranges from 0 to 1 or -1 depending on the strength of relationship.

8.1.2 When the value of $r = 1$, what does it mean? When it is negative 1 what does it mean?

Answer:
When the value of $r =1$, the correlation between the variable is direct and proportional such that the scatter diagram shows a straight line with a gradient of 1.

8.1.3 Explain what a negative correlation coefficient means

Answer:
When the correlation is –1 the scatter diagram and the line joining the best fit of the scatter is leaning from right to left. The values of the dependent or outcome variable becomes smaller as the values of the independent variable increases.
8.1.4 What does a correlation coefficient of zero mean

**Answer:**
A correlation coefficient of zero means that the relationship is rather circular and not linear.

8.1.5 How does sample size influence the strength of the correlation coefficient

**Answer:**
The larger the sample size the stronger the value of \( r \), or the stronger the credibility of the relationship shown. Therefore for a sample size of 1000, a correlation coefficient of 0.6 will be significant while for a sample of 20 the same value will not be statistically significant.

8.1.6 What are the assumptions behind a correlation?

**Answer:**
The assumptions are that the selection of the sample or samples was randomly done and the independent and dependent variables vary together in a joint distribution that is normally distributed.

8.1.7 Linear Regression

**Answer:**
Linear regression or is used to predict the value of the outcome variable (dependent) from values of the explanatory (independent) variable.

In simple regression, there is only one explanatory (independent) variable, which is used to predict the outcome.
In multiple regressions more than 1 independent variable is included in the prediction equation and the strengths of these in the whole may be calculated (defined).

8.1.8 Mention and explain the mathematical method used to calculate the statistical estimators in the regression equation.

**Answer:**
The mathematical method used to calculate the statistical estimators in the regression equation is the Least Square Method. It determines the equation of the line that provides the best fit to the points relating how the dependent variable changes with the independent variable.
Let \( a \) be the point where the line intercepts the \( Y \) axis and \( b \) the slope of the line. Then the equation of the line is \( Y = a + bX \)

*The slope indicates the direction of the relationship.* If the slope is positive, \( Y \) increases proportionally with \( X \). If the slope is negative, \( Y \) decreases as \( X \) increases.

*In the regression equation, for every value of \( X \), the value of \( Y \) can be predicted.* The difference between the actual value of \( Y \) and the predicted value of \( Y \) is \( E \). \( E \) therefore
provides a way of determining how fit the line fits the points. The **Least square method** uses this approach to find the line that minimizes this difference $E$.

![Diagram of a line with variables X and Y and the equation $r = y/x$](image)

8.1.9 What are the assumptions behind the equation

**Answer:**
We assume that for each value of variable $X$ the $Y$ variable has a normal distribution and the mean of the distribution is the value $\hat{Y}$. For whichever value of $X$ the standard deviation of $Y$ is the same. The assumption of this equal variation of $Y$ across the entire range of $X$ is called **Homogeneity**.
Learning Unit 9

Test of associations : risk, relative risk, odds ratio and attributable fraction

Learning Objectives

By the end of this unit you will be able to:

- Describe the risk, relative risk, odds ratio and attributable fraction
- Explain in which type of studies they are best applicable
- Calculate the risk, odds ratio, relative risk and attributable fraction
- Interpret the results of each of the calculations above

Cohort and Case Control Studies

A cohort study involves follow-up of individuals from the determination of exposure status to subsequent occurrences of disease. In contrast, a case-control study starts with the disease status of an individual and looks backward in time to determine the past exposure status to the risk factors of interest. For this reason, the terms prospective and retrospective are often used interchangeably with cohort and case-control respectively.

This unit focuses on the measures of association used to assess the strength of the relationship between a risk factor and the subsequent occurrence of disease.

In a cohort study a sample of individuals, some exposed to the risk factor of interest and some not, are followed over time, and the rates of subsequently contracting the disease in the two groups are compared.

9.1 Describe what is a relative risk, attributable risk, risk and odds ratio.

Answer:

Relative Risk (RR):

Relative Risk is a ratio and summarizes the strength of the association between the factor and the disease.

\[
\text{Relative Risk (RR)} = \frac{\text{incidence among exposed}}{\text{incidence among non-exposed}}
\]
A relative risk of 1 occurs when the incidences are the same in the two groups and is equivalent to no association between the risk factor and the disease. A relative risk greater than 1 occurs when the risk of disease is higher among those exposed to the factor than among the non-exposed. A relative risk less than 1 occurs when the risk is lower among those exposed, suggesting that the factor may be protective. An example is the reduced risk of diarrhoeal disease observed among infants that are breast-fed compared to those that are not.

The further the relative risk from 1, the stronger the association. It statistical significance can be tested using a 2x2 $\chi^2$ test.

Confidence interval for RR:
The formula given is that due to Miettinen. It is a test-based approximation calculated using the $\chi^2$ value found in the significance test, rather than a standard error.

$$95\% \text{ c.i.} = \frac{RR (1+/-1.96/x)}{RR}$$

Attributable risk (AR):
Attributable risk measures the magnitude of the excess in absolute terms.

$$\text{Attributable risk} = \text{incidence among exposed} - \text{incidence among non-exposed} = \frac{RR - 1}{RR}$$

In summary:
The relative risk is the best measure of the strength of an association between a risk factor and a disease. The larger its size, the more likely it is that the association is causal. Attributable risk, on the other hand gives, a better idea of the excess risk of disease experienced by an individual as the result of being exposed.

Risk (Incidence Risk):
There are two conceptually different ways of defining incidence; it maybe measured either as a risk or as a rate. The incidence risk is the probability that a person initially free from the disease develops it at some time during the period of observation. It is usually expressed as a percentage or, if small, as per 1000 persons.

$$\text{Risk (Incidence risk)} = \frac{\text{number of cases of disease in a specified period of time}}{\text{number at risk of contracting disease at beginning of period}}$$

The incidence rate on the other hand, is the rate of contracting the disease among those still at risk – when a person contracts the disease, they are no longer at risk.
Test of associations : risk, relative risk, odds ratio and attributable fraction

Incidence rate  =  \frac{\text{number of cases of disease in a specified period of time}}{\text{number of person years at risk during period (average number at risk of contracting disease during period x length of the period)}}

Odds Ratio (OR):

An alternative measure of incidence is the odds of disease to non-disease. This equals the total number of cases divided by those still at risk at the end of the study. The Odds Ratio is equal to the odds among the exposed divided by the odds among the non-exposed.

Using the table below:

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>Non - Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Non Exposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Odds ratio (OR)  =  \frac{a/b}{c/d} = \frac{ad}{bc}

The odds ratio may equivalently be considered as the ratio of the odds of exposure to non-exposure among the diseased (a/c) compared to the non-diseased (b/d). For this reason it plays an important role in case-control studies.

9.2 Explain the major applications of these tests of associations by answering the following:

- When is risk best applicable
- When is odds ratio best applicable
- When is relative risk best applicable
- When is attributable risk fraction best applicable

Risk is used in cohort studies, since it is possible to calculate the Incidence.

Odds ratio is used in a case control study or retrospective study where the risk is not possible to calculate but rather possible to calculate the odds ratio-between cases and controls.

Relative risk is best applicable when one wants to compare the probability of outcome between different levels of exposures.

Attributable risk fraction is an important information for public health and policy making as it provides the measure of the extent of the damage the risk has over a population.

9.3 Examples given to calculate.

9.4 Interpret the results of your calculations above: risk, relative risk, odds ratio and attributable fraction
Learning Objectives

By the end of this unit you will be able to:

- Describe the different types of error
- Explain how serious they may affect the study
- Describe the sources of error
- Provide practical examples of common errors and classify according to types

10.1 A team of researchers set out to conducting a longitudinal study of to determine the incidence of malaria among a cohort of children, who were followed up from birth to the age of one year. The study wished to also study the relationship of nutritional status and the incidence of the above parameters. It happened that for the first eight months, nobody remembered to calibrate the weighing machines which were used to monitor the growth weight gains during their regular attendance to MCH clinic and whenever they attended for treatment. It also happened that during active surveillance the temperatures of children were taken at different times of the day depending on the schedule of activities of the day. One of the data entry clerks had also difficulties differentiating between numbers 1 and 7, and numbers 3 and 5 handwritten by three of the field workers. It was noted that blood slide reading was intense and because there were only two technicians, they had to spend long hours, occasionally staying beyond the normal working hours to complete the work.

10.1.1 Identify the errors that were made during the above study and classify them into systematic and random errors.

Answer:

systematic errors from the non calibrated weighing machines. This is a grave mistake which can not be corrected if not discovered at the time of taking the measurements. The data entry clerk had the potential of introducing systematic errors in exchanging the numbers between 1 and 7, and 3 and 5. This can be identified and corrected by double entry and checking frequently for inconsistencies before establishing the clean data.

- random errors could be introduced by field workers through fatigue and the random variation of body temperature during the day.

The Slide readers could also introduce random errors through occasional loss of concentration or disturbances. It is important that such workers are given good space for work where there is no disturbances and they are not overworked.
10.1.2 Which of these errors are most serious and why?

Answer:
Systematic errors are more serious and lead to wrong conclusions (distorting the results) while random errors tend to cancel each other as they approximately happen with the same frequency in the comparison groups. Systematic errors are also difficult to detect and correct once they happen and all efforts must be made to prevent them.

10.1.3 What are the possible random and systematic errors that the blood slide readers could have committed.

Answer:
The possible systematic errors that the slide readers could have committed include the incorrect detection of *P.falciparum* asexual parasites. This could happen either by the inability of the reader to differentiate them from other malaria parasites or by artifacts introduced by using contaminated staining solution.

Random errors could arise from lack of concentration during slide reading due to occasional random disturbances or loss of concentration.

10.1.4 How could some of these errors be minimized during data entry?

Answer:
Double entry method is used to prevent both systematic and random errors that may be introduced by a data entry clerk. It is highly unlikely that both will make the same mistakes and therefore comparisons between two or more entries of the same data by different individuals will most likely reveal the mistakes.

10.1.5 What other types of errors could have been made by the field workers?

Answer:
Observer errors like correct reading of the temperature reading, changing the temperature taking source (axilla and rectal), and correct placement of the thermometer under the axilla so that it does not record the air temperature instead. -Respondent errors in enquiring about the child’s health status ( like fever within the past 24 hours), misunderstanding, faulty recalls, and likelihood of respondents giving the perceived correct answers.
Learning Unit 11

Budget and plan of action

Learning Objectives:

By the end of this unit you will be able to:

- Understand how to prepare plans and budget for operational research

11.1 Why should a study proposal contain a clear and justified budget?

Answer:
Any research study should contain a clear and justified budget for the following reasons:

- To ensure that all the intended activities will be conducted adequately
- To prevent the possibility of abandoning the study before its completion and so wasting time and money.
- To convince the fund provider that it is justified to fund the project at the requested level
- To create the basis for financial responsibility and accountability of the research team.

11.2 What are the most important elements to be included in a budget?

Answer:
All operational and administrative activities of the research work should be clearly spelt out and budgeted for. The important and basic elements of a research budget include:

- Personnel: All personnel to be involved in the study should be budgeted according to the amount of time they will spend for and on the research work (this includes the Principal Investigator/s, Investigators, technicians, other supportive staff)
- Equipment: Durable equipment
- Consumables: Include non durable equipment
- Travel: Local and International
- Computing Costs and Communication
- Administrative Costs or Institutional Overheads
- Other Operational Costs

All estimates should be based on real costs. If the project is lasting for several years consideration for exchange rate fluctuations and inflation should be made.

Under-budgeting is a grave problem which may lead to incompletion of the research in the manner and time intended. Budgets should be calculated most carefully to avoid this.

Tutor and Trainees to discuss the WHO budget requirements: Annex 1
11.3 Plan of Action

11.3.1 What are the main components of a plan of action?

Answer:
The main components of a plan of action are:
- A comprehensive listing of all activities to be implemented
- Indication of date of commencement for each activity
- Indication of expected date of completion of the activity
- Indication of responsible person/s for each activity
- Indication of important milestones for each activity

11.3.2 What is the best format of presenting the plan of action?

Answer:
Plans of action are best presented in the form of a chart having all the elements listed in 11.3.1

11.3.3 Develop a plan of action for your operational research proposal
Tutor to guide and assist participants.
Annex 1

WHO budget requirements
Summary of the Nuremberg code

1. Voluntary consent of the human subject is absolutely essential
   Legal capacity to give consent
   Free power of choice
   No force, fraud, deceit, duress or other constraint or coercion
   Sufficient knowledge and comprehension for an enlightened decision
   Personal duty and responsibility of researcher to obtain consent

2. Experiment should yield fruitful results for good of society
   Unprocurable by other means
   Not random or unnecessary

3. Experiment should be designed and based on animal experiment, knowledge of natural history of
disease or other problem under study
   Anticipated results will justify the experiment

4. Avoid unnecessary physical and mental suffering and injury

5. No experiment when a priori reason to believe death or disabling injury will occur

6. Degree of risk should never exceed that determined by the humanitarian importance of the
problem to be solved

7. Adequate facilities to protect study subjects

8. Conducted only by scientifically qualified persons

9. Subject has right to withdraw at any time

10. Scientist must be prepared to terminate experiment if continuation likely to cause injury,
disability or death.
Annex 3

Summary of the declaration of Helsinki

Drafted by World Medical Association, intended for medical doctors, but applicable to all working in the medical field.

Introduction

“The health of my patient will be my first consideration”
“The purpose of biomedical research involving human subject must be to improve diagnostic, therapeutic and prophylactic procedures and understanding of the aetiology and pathogenesis of disease”

I. Basic principles

1. Research must conform to accept scientific principles, be based on laboratory and animal experiments and thorough knowledge of the scientific literature.
2. Design and performance of experiment clearly formulated in an independently reviewed experimental protocol.
3. Conducted only by scientifically qualified person under supervision of a clinically competent medical person.
   Responsibility for human subjects rest with medical person, not with subjects of research
4. Importance of objectives is in proportion to inherent risk.
5. Experiment must be preceded by assessment of predictable risk in comparison to expect benefits.
   Concern for the interests of the subject must prevail over the interests of science or society
6. Rights of subjects to safeguard their integrity, and to privacy, must be respected
7. No experiment unless hazards predictable
8. Publish results accurately
9. Inform consent is mandatory; subject should understand aims, methods, benefits, hazards; should know they have the right to withdraw from the study.
10. When subject is in dependent relationship to the investigator, consent must be obtained by third party
11. Informed consent should be provided by a proxy when subject is legally incompetent
12. Research protocol should contain statement about compliance with the above ethical considerations.
II. Medical research combined with professional care (Clinical research)

1. Doctor must be free to use new diagnostic and therapeutic measures according to best judgement
2. Benefits, hazards, discomforts of new measures should be weighed against advantages of best current methods
3. Every patient should be assured of the best proven diagnostic and therapeutic method
4. Refusal of a patient to participate in a study must never interfere with the doctor-patient relationship
5. If the doctor considers it essential not to obtain informed consent, the specific reason should be stated in the protocol that is to be submitted for independent review
6. Medical research can be combined with professional care with the objective of acquiring new knowledge only to the extent that research is justified by its potential value to the patient

III. Non-therapeutic biomedical research involving human subjects (Non-clinical biomedical research)

1. It is the duty of the doctor to remain the protector of life and health of person on whom research is being conducted.
2. The subject should be volunteers
3. Investigators should discontinue research if in their judgment its continuation might be harmful to the subjects
4. The interests of science and society should never take precedence over the well-being of the subjects.
Annex 4

Summary of IEF guidelines

Part I

I. Obligations to subjects of research
   - Protect welfare
   - Obtain informed consent
   - Protect privacy
   - Maintain confidentiality

II. Obligations to society
    - Avoid conflicts of interest
    - Avoid partiality
    - Widen scope of epidemiology
    - Pursue responsibilities diligently
    - Maintain public confidence

III. Obligations to funders and employers
     - Specify obligations
     - Protect privileged information

IV. Obligations to colleagues
    - Report methods and results
    - Confront unacceptable conduct
    - Communicate ethical requirements

Part II

Commentary

This sets out the ethical framework of the guidelines and discusses each of the clauses in detail.
References

5. Kitua
12. Venning GR. Identification of Adverse reaction to new drugs......how were 18 important adverse reactions discovered and with what delays ? Br Med 1983;286:289-29
14. Uppsala Monitoring Centre. Safety Monitoring of medicinal products; Guideline for setting up and running a Pharmacovigilance Centre, 2000