Sileshi Lulseged. *Ethiop Med J, 2017, Vol. 55, No. 2*

TEACHING ARTICLE

CLINICAL EPIDEMIOLOGICAL RESEARCH METHODS: THE BASICS AND POTENTIAL PITFALLS

Sileshi Lulseged, MD, MMed*

ABSTRACT

Research is defined as a quest for knowledge through diligent investigation aimed at the discovery and interpretation of new knowledge using the scientific method. In the clinical setting, epidemiological methods are used to make a prediction about exposure to and/or health outcome for an individual based on scientific studies of groups of similar patients. Inadequacies in clinical epidemiological research methodologies constitute a challenge to generating and publishing sound, scientific clinical evidence. Flaws in methodology constitute the major reasons for this and indeed, for rejection of manuscripts submitted to health journals in general. Use of research methods not rigorous enough to answer a proposed research question and/or lack of adequate details on the research methods in the manuscripts submitted for publication are commonly observed features. These, not only compromise the quality of evidence generated and disseminated, but also contribute substantially to delays in the publication of manuscripts. This article describes the basic concepts of clinical epidemiological research and highlights common pitfalls and errors to be avoided.

INTRODUCTION

It has recently been emphasized that relevant and good quality evidence generated through well planned and carefully conducted clinical epidemiological studies are imperative to inform policy and clinical and public health services (1). In fact, it has long been recognized that well synthesized and properly disseminated scientific evidence is the linchpin of sound medical care (2-4). The best research evidence is usually found in clinically relevant research that has been conducted using sound clinical epidemiological methods. These methods apply epidemiological concepts in in clinical setting to make predictions about individual patients by counting clinical events in similar patients, using strong scientific methods for studies of groups of patients to ensure that the predictions are accurate (5). Indeed, health researchers need to work meticulously on the design and conduct of clinical epidemiological studies to ensure that the evidence they generate has the scientific rigor to positively influence clinical decision and policy.

This teaching article intends to present an overview of clinical epidemiological research methods with focus on the basic concepts and the methodological pitfalls that authors should guard against. It aims primarily to familiarize young scientists, researchers and health practitioners contributing to journal articles as authors and reviewers. The author hopes that the basics on clinical epidemiological studies described in this teaching article would serve as a motivation for investigators in the health field to give adequate attention to clinical epidemiological research methods in their research undertaking

HEALTH RESEARCH

The World Health Organization defines ***research*** as a quest for knowledge through diligent investigation aimed at the discovery and interpretation of new knowledge using scientific methods (6). In the health field, scientific research plays a very important role in the efforts to maintain health and combat diseases. Not only does it enable health care providers to properly diagnose and treat diseases, but also provides evidence for policies and decisions on health development. Available evidence suggests that not only investments in health research are generally not sufficient, but also that research is often not focused on priority health problems. Oftentimes it does not reflect best practices in terms of ethical standards and public accountability (7).

Health research can be classified in many different ways. A classification based on the field of research and specific focus area that is adopted from Fletcher and Fletcher (5) is given in Figure 1 below.

* School of Medicine, College of Health Sciences, Addis Ababa University
E-mail: sileshilulseged@gamil.com

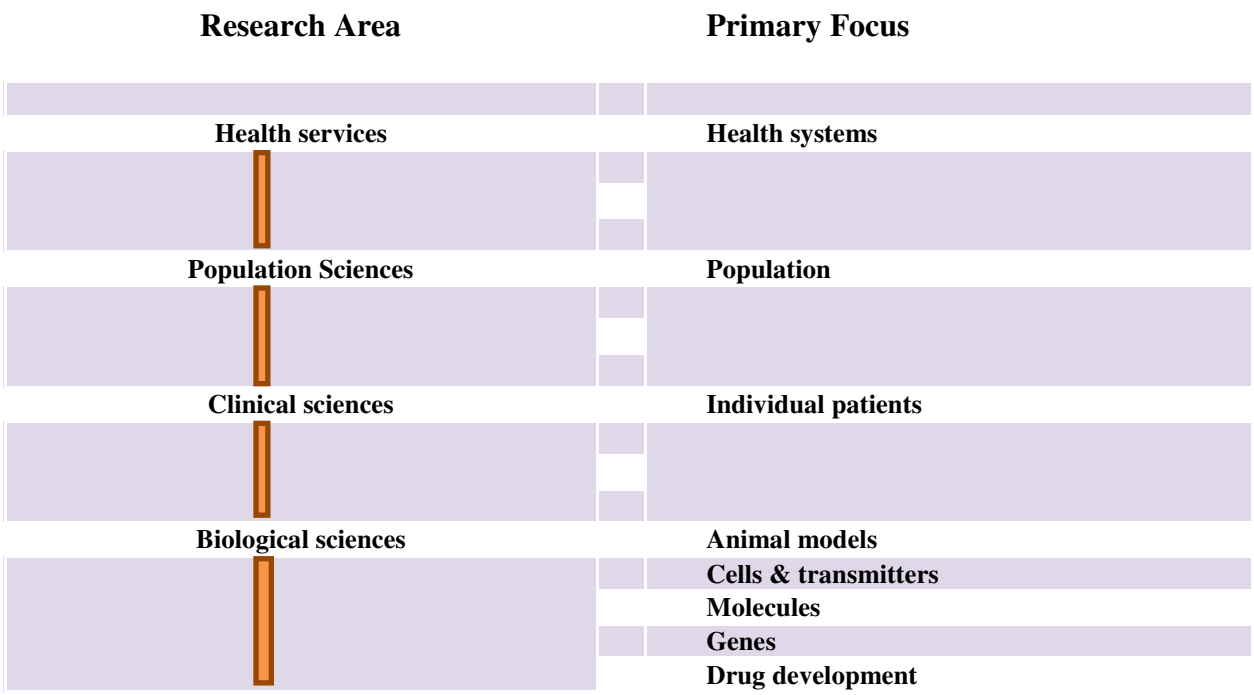

Figure 1. Classification and complementary relationships in health sciences research

CLINICAL EPIDEMIOLOGICAL RESEARCH

Research Question/Hypothesis:

The epidemiological research starts from questions raised at points of patient care on hospital wards or outpatient services outlets. One would need to articulate these questions into a well-formulated research question based on gap analysis and thorough understanding of the existing knowledge base. Irrespective of the field in which epidemiological research is being carried out there are some general guidelines for both the researcher and the user of research outputs (8).

- The research question (s) should be clearly identified and justified through critical appraisal of the current literature;
- A refutable or testable hypothesis may need to be generated and refined; and
- The hypothesis should be tested using an appropriate study design with valid and reliable measurements and appropriate analysis.
- The results should be interpreted with reference to the internal and external validity of the study

A hypothesis is a statement, at a higher and more abstract level, in which an attempt is made to generalize about the nature of the universe in which we live (8,9). Hypotheses are often stated in the ***null*** (H_1) form - no difference or no effect – as opposed to the positive/alternative form (H_1) - there is difference or effect. This emphasizes that scientific generalizations, while rarely verifiable, are generally falsifiable. This assumption remains true throughout the data collection and analysis steps of conducting the planned research until it is rejected or accepted based on the collected data.

Once a fairly well-defined research questioner hypothesis is formed, the best research method to address the research question/hypothesis is identified. The selection of the research method is core and is the single most important decision the investigator has to make in epidemiological research (10). Further considerations may include clinical setting, study design, selection criteria, data collection and analysis that are influenced by disease characteristics, prevalence estimates, resources and expertise available. As underscored by Suresh K, et al., (11) not spending adequate time by investigators to consider all these aspects in the planning phase of the study is a common pitfall.

Data collection methods: Primary or secondary data can be collected using quantitative or qualitative methods or a mix of the two methods. Predefined objectivity is very important in quantitative research (11,12). Also the issues of clustering of respondents in some form of communal units and weighting of sample estimates when appropriate are extremely important for the validity of findings. These are not usually well addressed by the contributors to health journals. In contrast, qualitative research involves the recording and analysis of data in an attempt to uncover the deeper meaning and significance of human behavior and experience, including contradictory beliefs, behaviors and emotions. The method is mostly used to answer the question why. Data is collected in textual form on the basis of observation and interaction with study participants, e.g. through participant observation, in-depth interviews and focus group discussion.

The differences between the two methods, adapted from Suresh K, et al. (11), are summarized in Table 1.

Table 1. Main differences between qualitative and quantitative research methods

Characteristic	Quantitative research	Qualitative research
Type of data	Phenomena are described numerically	Phenomena are described in a narrative fashion
Analysis	Descriptive and inferential statistics	Identification of major schemes
Scope of inquiry	Specific questions or hypotheses	Broad, thematic concerns
Major advantage	Large sample, statistical validity, accurately reflects the population	Rich, in-depth, narrative description of sample
Major disadvantage	Superficial understanding of participants' thoughts and feelings	Small sample, not generalizable to the population at large

A pragmatic approach to epidemiological research involves using the method which appears best suited to the research question/problem. This approach often uses any of the methods, techniques and procedures typically associated with quantitative or qualitative research. It recognizes that each method has its limitations. Sometimes it is possible to transform qualitative data into quantitative data. Mixing the two methods has the advantage of enabling ***triangulation***, a common feature of mixed methods studies (16).

KEY ISSUES IN QUANTITATIVE RESEARCH DESIGN

Sampling methods and Sample size:

Samplingis a principle that specifies the conditions and guides the process of selecting members of the population to participate in the study. The choice of sampling method determines the accuracy of research findings, reliability and validity of the study and has immense implications on the overall quality of the study. To ensure reliable and valid inferences from a sample, probability sampling techniques are preferred choices because they are believed to generate unbiased results. Probabilistic sampling techniques include:

- ***Simple random sampling*** - every subject has an equal chance of being selected for the study by using a table of random numbers or a computer-generated list of random numbers.
- ***Systematic random sampling*** - a method in which every k^{th} item is selected; k is determined by dividing the number of items in the sampling frame by sample size and the starting point from the first group is selected using simple random sampling.
- ***Stratified random sampling*** - a method in which the population is first divided into relevant non-overlapping strata or subgroups and then, using the simple random sample method, a sample of predetermined size is drawn from each stratum.
- ***Cluster sampling -*** dividing the population into non-overlapping clusters and selecting some clusters randomly. It is also possible to have more layers of clustering which is called ***multi-stage*** cluster sampling.

Sample size: The most important question that a researcher should ask when planning a quantitative study is "How large a sample do I need?" Variables such as prevalence, expected confidence level and expected treatment effect need to be predetermined in order to calculate sample size. If the sample size is too small, even a well-conducted study may fail to an-

swer its research question or detect important effects or associations, or may estimate those effects or associations too imprecisely. Careful consideration of optimal sample size and ***statistical power*** analysis during the planning and design stages of clinical research is crucial. The power of a binary hypothesis test is the probability that the test correctly rejects the null hypothesis (H_0) when the alternative hypothesis (H_1) is true (17). Statistical power is directly related to sample size; the bigger a sample, the higher the power to detect an effect. Investigators could possibly miss a real effect (Type II error) simply because of underpowered studies, an uncommon pitfall in clinical research.

Bias and Confounding: Most clinical epidemiological studies try to determine if an association exists between a risk factor (exposure) and a disease (out-come). In appraising the observed relationship, it is important to keep in mind that the association may arise from either chance, bias or confounding or a combination of these. The role of 'chance' is dealt with by calculating probability values of study results being true or due to chance alone, i.e., through the use of '***p values***' and '***confidence intervals***'. Bias and confounding also affect observations and lead to false conclusions regarding association between an exposure and outcome (15,16). The effect of bias and confounding cannot be eliminated entirely, but can be minimized.

Bias: Bias may be defined as any systematic error in an epidemiological study that results in an incorrect estimate of the association between exposure and disease/outcome. It results due to a systematic error introduced into a study during the design phase (sampling, subject recruitment), or the data collection phase (measuring exposure and outcome). There are many types of bias encountered in epidemiological research. Sampling bias, selection bias, observation bias, misclassification bias and losses to follow-up constitute examples of commonly occurring biases. A detailed list and analysis of the different types of biases is given by Sacket D. (17) and Althubaiti A. (18).

Confounding: Confounding has traditionally been defined as bias arising from the co-occurrence or mixing of effects of extraneous factors, referred to as confounders, with the main effect(s) of interest. A confounding factor (confounder), also known as a third variable or a mediator variable, can adversely affect the relation between the exposure variable and outcome variable (19). Confounding factors, if not controlled for, cause bias in the estimating the impact of the exposure being studied. The effects of confounding can result either in an observed difference between study populations when no real difference exists or no observed difference between study populations when a true association does exist. Mixing up confounding with **effect modification** and **interaction** is a common error. Effect modification is all about stratification and occurs when an exposure has a different effect among different subgroups and, unlike confounding, it is associated with the outcome but not the exposure; interaction is the interdependent operation of two or more biological causes to produce, prevent or control an effect (20).

Measurement issues: Whenever a test or other measuring device (e.g. questionnaire) is used as part of the data collection process, the validity (lack of systematic error) and Precision (lack of random error) of that test is important (21). In research we are relying on the results to show support or a lack of support for our hypothesis. If the data collection methods are erroneous, the data we analyze will also be erroneous. Thus, the validity and precision together determine the accuracy of the measurement, which is essential to make valid statistical inference from an epidemiological clinical research (22,23).No statistical method will have enough capacity to correct erroneously collected data.

Validity:

Measurement Validity: defined as "an expression of the degree to which a measurement measures what it purports to measure". Occasionally we have access to the 'gold standard' which is believed to provide a measure of some characteristic that is close to the truth. The measure used in a study is compared to a gold standard to assess its performance (Table 2).

Table 2: Performance of Study test result

Study Measurement	**Gold Standard**		
	Condition Present	Condition not Present	
Condition Present	a	b	a+b
Condition Not Present	c	d	c+d
	a+c	b+d	

Sensitivity = $\frac{a}{a+c}$ **Specificity** = $\frac{d}{b+d}$ **Positive Predictive Value** = $\frac{a}{a+b}$ **Negative Predictive Value** = $\frac{c}{c+d}$

Study Validity: defined as the degree to which the inference drawn from a study are warranted when account is taken of the study methods, the representativeness of the study sample, and the nature of the population it is drawn from.

- ***Internal validity*** dictates how an experimental design is structured and encompasses all of the steps of the scientific research method. Even if your results are great, sloppy and inconsistent design will compromise your integrity in the eyes of the scientific community.

- ***External validity*** is the extent to which the results of study can be generalized to other situations or the target population.

Reliability**:** Reliability ***(Precision)*** is the degree to which an assessment tool produces stable and consistent results. It refers to the extent to which the research measure is a consistent and dependable indicator of the investigation. A reliability coefficient is often the statistic of choice in determining the reliability of a test (23). This coefficient merely represents a correlation, which measures the intensity and direction of a relationship between two or more variables.

STATISTICAL ANALYSIS

Statistical analysis is an essential technique that enables a medical researcher to draw meaningful inference from a given data. Improper application of study design and data analysis may render insufficient and improper results and conclusion. Converting a medical problem into a statistical hypothesis with appropriate methodological and logical design and then back-translating the statistical results into relevant medical knowledge is a real challenge (22).

Descriptive Statistics: Descriptive statistics are a means of organizing and summarizing observations in a sensible way. Together with simple graphic summaries, they form the basis of virtually every quantitative analysis of data (24). In comparison, with inferential statistics, we are trying to reach conclusions that extend beyond the immediate data alone. For instance, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance alone.

Univariate Analysis**:** Univariate analysis involves examination across cases of one variable at a time. There are three major characteristics of a single variable that we tend to look at the distribution; the central tendency; and the dispersion of data. In most situations, we would describe all three of these characteristics for each of the variables in the study.

Distribution**:** is a summary of the frequency of individual values or ranges of values for a variable. The simplest distribution would list every value of a variable and the number of persons who had each value. One of the most common ways to describe a single variable is with a ***frequency distribution***. Frequency distributions can be depicted in two ways, as a table or as a graph. Distributions may also be displayed using percentages.

Central Tendency: The central tendency of a distribution is an estimate of the "center" of the distribution of value, estimated by the ***Mean*** (used only when the distribution of the values of a continues variable are normally distributed), ***Median*** (used when distribution is not normal; it is the score found at the exact middle of the set of values, which are arranged from the smallest to the largest in a sequence), and ***Mode*** (the most frequently occurring value in the set of scores).

Dispersion**:** Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation. The ***range*** is simply the highest value minus the lowest value. The ***standard deviation*** (used with the ***mean***) measures the degree of spread of individual values in reference to their mean and is a more accurate estimate of dispersion because an outlier can greatly exaggerate the ***range***. ***Interquartile range (IQR),*** also called the midspread or middle 50% in the rank ordered set (24),is the appropriate measure of dispersion when the ***median*** is used as a measure of central tendency. Assuming that the distribution of scores of a variable is normal (approximately normal) or bell-shaped, where ***mean*** is used as a measure of central tendency, approximately 68% of the scores in the sample fall within one standard deviation, 95% of the scores within two standard deviations and 99% within three standard deviations of the mean. Use of a measure of dispersion not matching with the correct measure of central tendency a common methodological error seen in manuscripts submitted for publication.

Tables and Graphs: Tables and graphs are an excellent way to display data in an organized fashion. By putting data in tables one can easily set up a graph to illustrate the data.

Tables: The summary table is a visualization that summarizes statistical information about data in table form

Graphs*: *Bar graphs/Bar charts are excellent ways to display a frequency distribution for nominal or ordinal data. Bar graphs are useful to get an overall idea of trends in responses - which categories get many versus few responses.

Histograms are the most commonly used graphs to show frequency distribution (shape) of a set of continuous data set.

Line graphs are most useful in displaying data or information that changes continuously over time.

Pie graphs are used when considering how many parts of a whole there are.

Other graphs commonly used to summarize data include ***frequency polygons***, ***one-way scatter plots***, ***box plot*** and ***two-way scatter plot*** (26).

STATISTICAL SIGNIFICANCE:

In statistical hypothesis testing, the "***significance level (α)"*** is the probability of rejecting the null hypothesis, given that it is true, but no enough evidence in the data to support it. It is a misleading term that many researchers do not fully understand (25). The term ***significance*** does not imply ***importance*** and the term ***statistical significance*** is not the same as research, theoretical, or practical significance. A ***statistical significance*** (or a ***statistically significant*** result) is attained whenever the observed ***p-value*** of a ***test statistic*** is less than the significance level defined for the study. When testing hypotheses, there are two types of error as shown in Table 3 below:

Table 3. Types of error in hypothesis testing

	Accept H_0	**Reject H_0**
H_0 True	**Correct Decision**	**Type I Error (alpha; α)**
H_0 False	**Type II Error (beta; β)**	**Correct Decision**

The *p*-value is the probability of obtaining results at least as extreme as those observed, given that the null hypothesis is true (a ***type I error***). If the *p*-value of an observed effect is less than the significance level, an investigator may conclude that the effect reflects the characteristics of the whole population, thereby rejecting the null hypothesis. A significance level is chosen before data collection, and typically set to 5% (26) and used as one of the inputs in sample size estimation. In a two-tailed test, the rejection region for a significance level of $\alpha=0.05$ is partitioned to both ends of the sampling distribution and makes up 5% of the area under the curve.

Effect size: A statistically significant result may have a weak effect. To gauge the research significance of their results, researchers are encouraged to always report an effect size along with *p*-values. There are situations where researches report equivalence of comparison groups rather than statistical significance of the effect size.

ETHICAL CONSIDERATIONS:

The World Medical Association Declaration of Helsinki emphasizes that it the duty of physicians who are involved in medical research to protect the life, health, dignity, integrity, right to self-determination, privacy, and confidentiality of personal information of research subjects (27). The Ethiopian national health research guideline (28) re-affirms that health research should strictly observe these ethical principles. The World Health Organization (WHO) also recommends that health research must be responsive to the health needs or priorities of the communities where the research is conducted (29). Researchers also need to recognize, that there is a strong relationship between ethics and research methods. Medical research involving human subjects must conform to the generally accepted scientific principles, and be based on a thorough knowledge of the scientific literature and other relevant sources of information (30).

CLINICAL EPIDEMIOLOGICAL RESEARCH METHODS

From an epidemiological point of view the type of research method is very important since there are a number of advantages and disadvantages to various types, and, in particular, there are traps and pitfalls one should seriously guard against. The different types of studies that fall under experimental and non-experimental epidemiological research methods are summarized in Table 4.

Table 4: Types of Research Studies

Non-Experimental (Observational) Studies:		
Cross-sectional studies/Surveys Case series/Case reports Ecological (Correlational) studies	}	Descriptive studies
Case-control studies Cohort studies		
Experimental studies:		
Clinical trials Field trials		
Studies to review scientific evidence Systematic reviews Meta-analyses		

The types of studies are shown in Figure 2 below in terms of the rigor of evidence they produce.

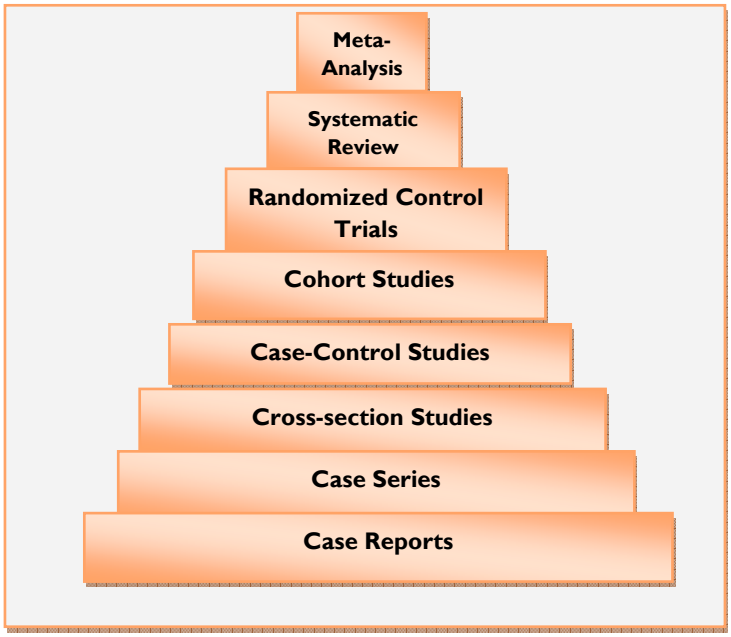

Figure 2: Research study methods with rigor of evidence increasing from the base upwards to the apex of the pyramid

The decision as to which study design should be used is a function of the research question being asked and the logistics of the study involved including time, cost and other constraints. In the following sections the most important distinguishing features of the various study methods are described.

I. NON-EXPERIMENTAL (OBSERVATIONAL) STUDIES

DESCRIPTIVE STUDIES:

In contrast to analytical or experimental studies, descriptive studies are not structured formally and do not aim at specifically testing hypotheses (31). Descriptive studies provide researchers the first clues in identifying an association between a given exposure and a disease. They provide a wealth of material which allows the generation of hypotheses, which can then be tested by analytical or experimental designs. Both qualitative and quantitative data collection techniques may be used in data collection.

CROSS-SECTIONAL STUDIES:

A cross-sectional study or community (population) survey entails the collection of data on a cross-section of the population or a proportion (sample) of it. The most important distinguishing feature about a cross-sectional study is that both 'exposure' and 'outcome' are measured at the same time. Information is obtained from the study subjects only ones and provides a 'snapshot' of the health (or other) experience of the population at one point in time (16). A cross-sectional study provides prevalence at a particular point in time (point prevalence) or over a period of time (period prevalence). They are more appropriate for measuring the relationship between fairly permanent characteristics in individuals and chronic diseases or stable conditions.

A cross-section study usually does not aim at testing a hypothesis about an association, and is thus descriptive in nature. A ***descriptive cross-sectional study*** usually aims at estimating the magnitude of a specified problem in a defined target population. Occasionally a cross-sectional study may be analytical in nature, such as when the investigator measures exposure and outcome factors as they exit at one point in time in a representative sample of the population. An ***analytical cross-sectional study*** is only valid when the current values of exposure factor are extremely stable over time. The advantage and disadvantages of the method is shown in Table 5.

Table 5: Advantages and disadvantages of cross-sectional studies

Advantages	**Disadvantages**
• Relatively quick and inexpensive • Researcher has control over the selection of study subjects • Researcher has control over the measurement used • Allow studying several factors and outcomes at one time • Less prone to recall bias among the subjects • Ethically concerns not more than minimal • Provides early clues to hypothesis generation • Provide Prevalence information • Allow a risk statement to be made, although thesis not precise • Often a good starting point for a cohort study • Provide a wealth of data for health systems research	• Not possible to establish temporality between exposure and outcome • Establish association at most, not causality • Provide no direct estimate of risk • Potential bias in sampling and/or bias from selective survival (Neyman bias) • Potential bias in measuring exposure • Confounders may be unequally distributed • Group sizes may be unequal • Not feasible for rare conditions • Does not yield incidence or true relative risk

The design and conduct of cross-sectional studies:

A schematic presentation of cross-sectional studies is given in Figure 3. They usually start with a reference population, from which a random sample is taken.

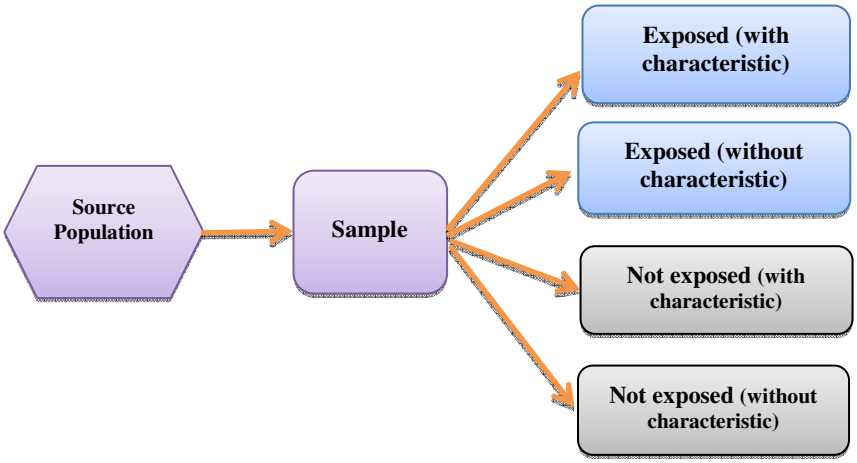

Figure 3: Design of a cross-sectional study

Sampling: For extrapolation from the study data, the sample of study subjects should be representative of the larger population from which they were obtained and need to have been selected in an unbiased manner. If these conditions are met then the study is deemed to have good ***external validity (generalizability)***. Some of the issues that need to be considered in obtaining study samples are given in Figure 4 below.

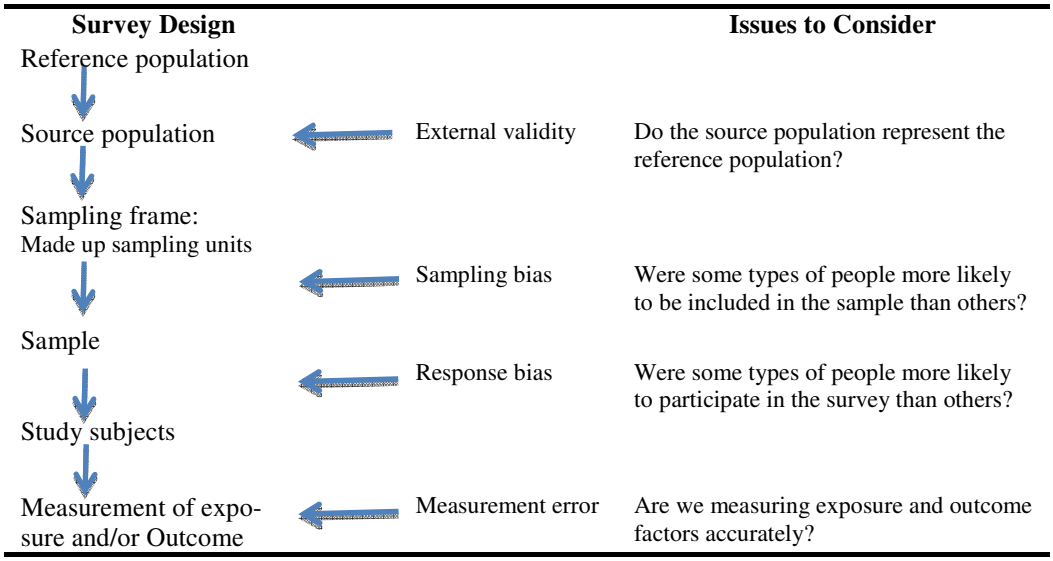

Figure 4. Issues to be considered in the sampling process

Measurement issues: The basic technique in a cross-sectional study is to use a standardized measure for the exposure and outcome of interest. In the measurement of a disease outcome, case definition is very important. The evaluation of method involves two kinds of assessment – validity and reliability.

Analysis and interpretation issues: Prevalence, the main outcome measure obtained from a cross-sectional study, is estimated by:

$$\text{Prevalence} = \frac{\text{No. cases in defined population at one point in time}}{\text{No. persons in defined population at same point in time}}$$

In analytical cross-sectional studies, the ***odds ratio (OR)*** can be used to assess the strength of an association between a risk factor and the outcome of interest, provided that the current exposure accurately reflects past exposure (32).

CASE SERIES/CASE REPORTS

Case series and case reports describe the experience of a single patient or a group of patients with a similar diagnosis and constitute an important interface between clinical medicine and epidemiology (33). From case series, one may calculate proportional distribution, which consists simply of percentages of the total number of cases that belong to a specific category of age, sex, ethnic group or other characteristic.

1.3 ECOLOGICAL DESCRIPTIVE STUDIES

Ecological (correlational) studies are descriptive studies on health or other outcomes that use aggregates as the unit of observation (e.g. family, clan or school) or an ecological unit (a village, town or country) defined either geographically or temporally (8,16). The association between the exposure and outcome factor across populations is measured by the correlation coefficient (r) which quantifies the extent to which there is a linear relationship between exposure and disease. They value of the correlation coefficient varies between -1 (strong inverse relationship) to +1 (strong positive relationship). The most serious flaw in correlational studies is the risk of ecological fallacy, when the characteristics of the geographical unit are incorrectly attributed to the individuals. The advantage and disadvantages of the method is shown in Table 6.

Table 6: Advantages and disadvantages of Ecological/correlational studies

Advantages	Disadvantages
▪ Quick and inexpensive ▪ Use information already available ▪ Can be used as the first in exploring the relationship between an exposure and disease ▪ Data can be used from populations with widely differing characteristics	▪ A link to individuals cannot be made ▪ Potential confounding factors cannot be controlled ▪ A lack of correlation may not mean lack of association between exposure and disease ▪ The 'average' exposure level is assessed; may mask dose-response relationships

CASE-CONTROL STUDIES:

Case-control studies constitute the simplest and most commonly used analytical strategy in clinical research. Designed and executed in skilled hands they are capable of providing important insights into the etiology of many conditions. The problem with case-control studies is possible that they seem so easy to do, a misconception which may lead to fundamental errors in the planning, implementation of the study and analysis of results.

The design is backward-looking examining exposure histories of cases and controls (18,34). As illustrated in Figure 5, the study subjects are selected on whether or not they have the disease (or other outcome) of interest. Cases (those with disease) are then compared to controls (those without the disease) in terms of their history of exposure.

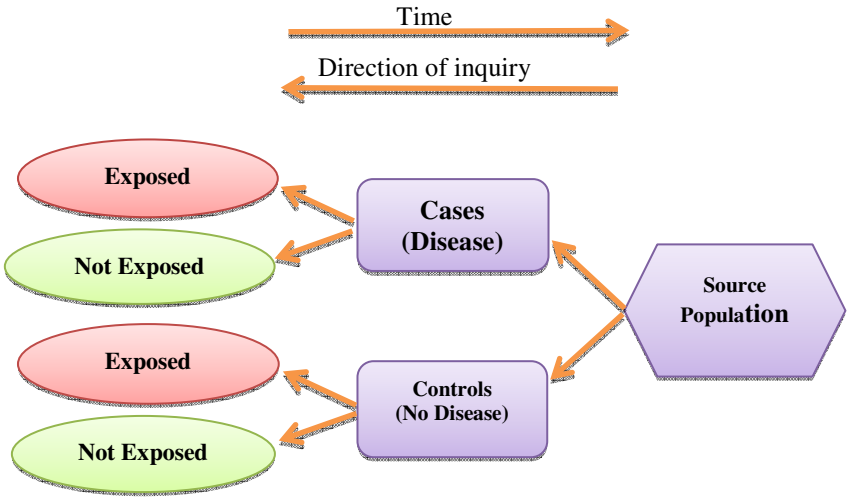

Figure 5: Design of a case-control study

The selection of controls may involve matching by certain characteristics (e.g. age, sex, race and socioeconomic status), factors other than the risk factor being investigated (35). Of note, the availability in recent years of multivariate statistical techniques which can control for the effect of multiple potential confounders has resulted in a reduced reliance on matching. The advantages and disadvantages of the case-controlled study design is shown in Table 7 and calculation of the OR given in Table 8.

Table 7: Advantages and disadvantages of Case-control studies

Advantages	Disadvantages
• quick and cheap • only feasible method for very rare disorders or those with long lag between exposure and outcome • fewer subjects needed than cross-sectional studies • little problem with attrition • could be the earliest practical observational strategy for determining an association • possible to demonstrate a dose-response or gradient relationship • possibility for a hybrid design (nested in a cohort study); a powerful strategy	• reliance on recall or records to determine exposure status • selection of control groups is difficult • potential bias: recall, selection, misclassification • absence of epidemiological denominators (population at risk) – calculation of incidence rates/ attributable risks, impossible • temporality is a serious problem • incapable of disclosing other conditions related to the risk factor • mostly performed in hospitals- liable to Berkson's fallacy or the effect of differing admission rates

Table 8: Calculation of the odds ratio

RESULT		Outcome	
		Present	Absent
Exposure	Yes	a	b
	No	c	d

Direction of Sampling

$$OR = \frac{a \times d}{b \times c}$$

COHORT STUDIES

A cohort study is an observational longitudinal research method which begins when a group of people (a cohort) initially free of the disease, are classified according to a given exposure (exposed group and unexposed group), and then followed up over time. It is the best way to identify incidence and the natural history of a disease (36). Both groups are then observed over a specified period to find out the risk each group has to develop the outcome(s) of interest. This is illustrated diagrammatically in Figure 6.

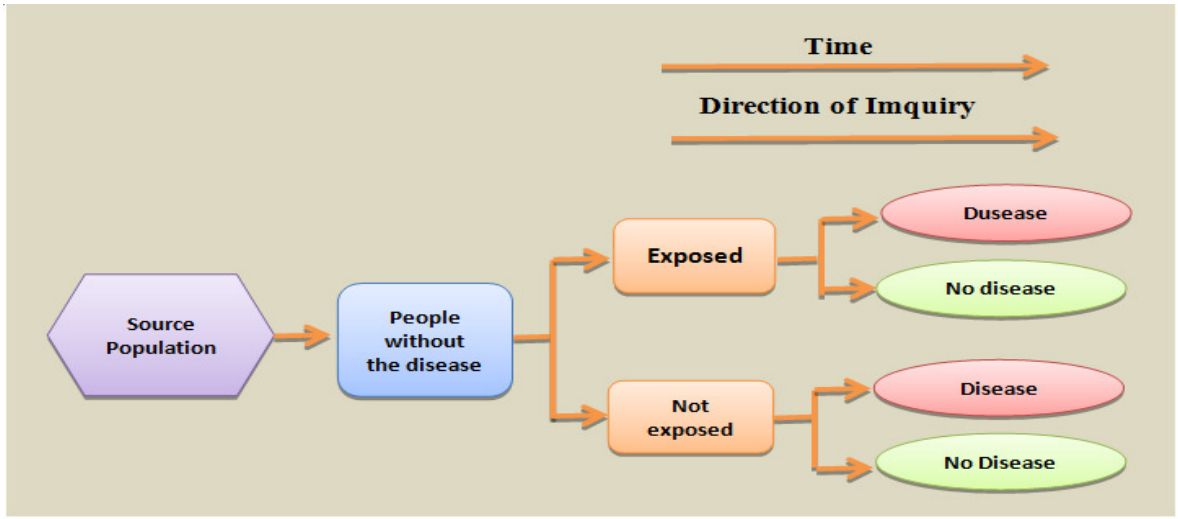

Figure 6: Design of a cohort study

i. PROSPECTIVE COHORT STUDY

In a prospective cohort study the starting time for observations (time zero) is now, and the study population is followed into the future. The exposure of interest may or may not have occurred when the study was initiated, but the outcome has not (16). The advantages and disadvantages of this design is given in Table 9.

Table 9: Advantages and disadvantages of prospective cohort studies

Advantages	Disadvantages
• ethically safe • allows the temporal sequence between exposure and disease and to be clearly established • can be used to measure multiple exposure and outcome factors • suitable to measure the effect of rare exposures • subjects can be matched • can establish timing and directionality of events • eligibility criteria and outcome assessments can be standardized • allow the possibility of measuring directly relative and attributable risks • less chance for recall or survival bias • administratively easier and cheaper than RCT	• controls may be difficult to identify • exposure may be linked to a hidden confounder • blinding is difficult • sample sizes required for cohort studies are extremely large • Costly in time, personnel, space and patient follow-up • randomization not present/ required • relatively inefficient for studying rare outcomes (large sample sizes or long follow-up necessary) • participants' attrition can be as serious problem (if beyond 10-15%, serious potential bias) • investigators' attrition may also be a problem • changes may occur in the environment (may confuse the issue of association & attributable risk) • procedures may influence behavior of persons overtime influencing outcome development (Hawthorne effect)

Design and data collection:

The sample comprising both the exposed and non-exposed study subjects can come from a variety of sources, including:

- a community cohort of specific age and sex
- Population based birth cohort
- an exposure cohort, e.g. radiologists, smokers, users of oral contraceptives
- a birth cohort, e.g. school entrants
- an occupational cohort, e.g. miners, military personnel
- a marriage cohort; and
- a diagnosed or treated cohort, e.g. cases treated with radiotherapy, hormonal treatment

Scientific and logistic considerations may include frequency of the exposure of interest in a particular subgroup of the population, the need to obtain complete and accurate information on exposure and the likely success of being able to follow-up particular study subjects over a long period of time. The groups being compared should be as similar as possible with respect to all other factors except for the exposure of interest. Data should be collected over the follow-up period as this will have implication on the quality and interpretation of the findings.

The data to be collected includes:

- data on the exposure of interest;
- data on the outcome of interest. and
- characteristics of the cohort that might confound the association under study.

The nature of the exposure of interest needs to be carefully and precisely defined and accurately assessed. Ascertainment of outcome is a crucial component of a cohort study. The goal is to obtain complete, comparable and unbiased information about the health experience of every subject from the start of the study to the designated time of follow-up. Person(s) extracting the information from various sources should ideally be blinded and that the same procedures and levels of effort are expended in ascertaining outcomes for both the exposed and non-exposed individuals.

Data collection in a prospective cohort study may involve:

- interview surveys with follow-up procedures
- medical records monitored over time
- medical examinations and laboratory testing
- record linkage of sets with exposure data and sets with outcome data

Analysis and interpretation of results:

Incidence: Incidence (absolute risk) is a measure of disease that allows us to determine a person's probability of being diagnosed with a disease (number of newly diagnosed cases of a disease) during a given period of time. An ***incidence rate*** is the number of new cases of a disease divided by the number of persons at risk for the disease. Persons at risk can also be changed into person time that a hypothetical person was exposed.

Relative Risk: Relative risk (RR) is the ratio of the incidence of disease (outcome) among the exposed relative to the incidence of disease (outcome) among the non-exposed. The formula used to calculate RR given in Table 10.

Table 10: Calculation of the Relative Risk in Cohort Studies

RESULT		Outcome	
		Present	Absent
Exposure	Yes	a	b
	No	c	d

Direction of Study

$$\mathbf{RR} = \frac{\mathbf{a}}{a+b} \div \frac{\mathbf{c}}{c+d}$$

Cumulative incidence: This index of disease frequency is the ratio of the number of new cases or events in a specified period of observation to the total population at risk during that particular period. This incidence measure provides an estimate of the probability or risk of developing disease/outcome among all members enrolled in the study.

Incidence density (person-time approach): This approach is an improvement over the conventional measure of incidence, because it takes into consideration both the number observed and the duration of observation for each individual. Thus, if 30 individuals were observed as follows: 10 for two years, 5 for three years, and 15 for four years, they would contribute (10x2)+(5x3)+(15x4) = 95 person-years of observation, which would become the denominator. The numerator is the number of new cases observed in these groups over the specified period of time. This gives an incidence rate per person-year, called the incidence density.

Other measures: other basic measures of effect used in cohort studies (36):
Population attributable risk percent (ARP%); and
Etiologic fraction (EF).

There are a number of biases that need to be considered in interpreting the results of cohort studies, including:
Selection bias (in retrospective cohort in particular);
Ascertainment bias (both exposure an outcome);
Misclassification bias (with respect to exposure status);
Participation bias (some characteristics of those who agree to participate may be different from those who do not); and
Loss to follow-up (differential loss of over 10-15%).

Nested case-control studies: A nested case-control study is an offshoot of a prospective cohort study (37). These studies use the approach of a case-control study. However, they use data gathered as part of a cohort study. They are less prone to bias than standard case-control studies. All things being equal, the strength of nested case-control data falls somewhere between that of standard case-control studies and cohort studies.

ii. RETROSPECTIVE (HISTORICAL) COHORT STUDIES

Retrospective cohort (or historical cohort) follows the same direction of inquiry as a prospective cohort study illustrated in Figure 4 above. However, this study design uses information that has been collected in the past and kept in files or databases. Patients are identified for exposure or non-exposures and the data are followed forward to an effect or outcome of interest to the present. A historical cohort study depends upon the availability of data or records that allow reconstruction of the exposure of cohorts to a suspected risk factor and follow-up of mortality or morbidity over time. Table 11 shows that advantages and disadvantages of a retrospective cohort design.

Table 11: Advantages and disadvantages of retrospective cohort studies

Advantages	Disadvantages
• share several advantages of the prospective cohort • all requirements satisfied, a historical cohort may suffer less from the disadvantages of time and expense.	• All of the relevant variables may not be available • It may be difficult to ascertain that the study population was free from the condition at the start of the comparison • Attrition problems may be serious due to loss of records, incomplete records or difficulties in tracing or locating all study population • require ingenuity in identifying suitable populations and in obtaining reliable information concerning exposure and other relevant factors

Note: Issues related to analysis and interpretation of results as well as ethical aspects of retrospective cohort studies are similar to prospective cohort studies describe above.

Ethical considerations: A serious ethical problem may arise in cohort studies when it becomes apparent that the exposed population is manifesting significant disease excess before the follow-up period is completed. Care must be taken to ensure that a cohort design satisfies the requirements of epidemiological research, particularly with regard to appropriate sampling, construction of comparison groups, handling of missing data, application of appropriate statistical methods and other prerequisites.

II. EXPERIMENTAL STUDIES (CLINICAL TRIALS/INTERVENTIONS)

RANDOMIZED CONTROLLED CLINICAL TRIALS

Definition and basic structure: Randomized controlled trials (RCTs) are carefully planned experiments that introduce therapeutic or other interventions (preventive, services) to study their efficacy (as compared to effectiveness in field trials) on real patients. RCTs are considered as the 'gold' standard' of all epidemiological clinical research studies (38). When done properly they provide the level of assurance about the validity of the study results that cannot be matched with any other study design. In order to ensure the validity of RCTs, a number of issues need to be considered, including the selection of study population, allocation to treatment (other interventions) and control groups, the maintenance and assessment of compliance and accurate assessment of outcomes. In the selection of a study population, it is important to guard against possible biases, consider that sufficient number of study subjects (statistical power) will be obtained and that a sufficient number of endpoints (outcomes of interest) will conceivably be achievable. It is also important to consider the logistics of obtaining complete and accurate follow-up information on the study subjects for the duration of the study. The advantages and disadvantages of RCTs are given in Table 12.

Table 12. Advantages and disadvantages of Randomized Controlled Trials

Advantages	*Disadvantages*
• treatment and control group very comparable • unbiased distribution of known and unknown confounders • provides best evidence for cause and effect • blinding more likely • allows use of statistical analysis with few inbuilt assumptions	• expensive in terms of time and money • may be unsuitable due to problems of cooperation and rarity of outcome • tend to induce artificial situation such as: patients involved may not be representative of patients usually in clinical practice; highly standardized interventions may differ from usual clinical practice (efficiency vs. effectiveness; volunteer bias • incurs more than minimal risk

There are two types of designs used in RCTs:

i. Parallel design

In parallel design, more than one arm is employed. One arm (investigational arm) gets the new treatment/intervention and the other arm (control arm) is given either a placebo, or a conventional treatment/intervention.
The basic structure of a parallel design for RCT is given in Figure 7.

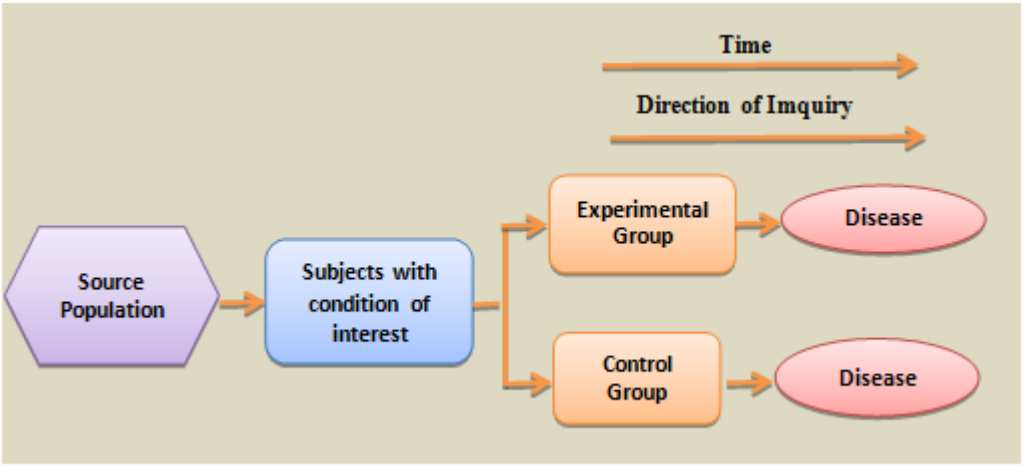

Figure 7 : Basic structure of parallel randomized controlled trial

The flow in the selection and allocation of patients to the arms of the RCT is given in Figure 8 below.

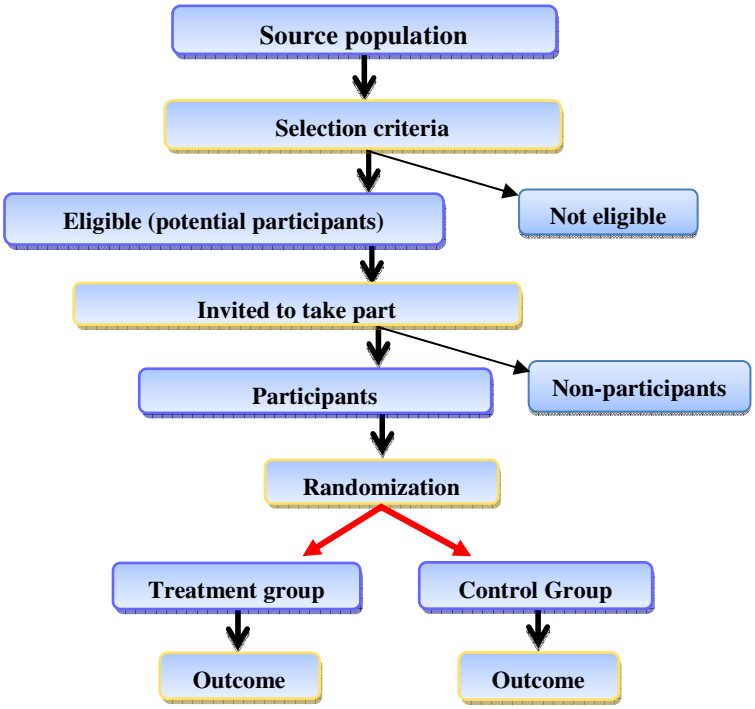

Figure 8 Participants selection and allocation in RCT

ii. Crossover randomized Controlled trials:

A **crossover trial** is a longitudinal study in which subjects receive a sequence of different treatments (or exposures). This type of trial involves a repeated measures design in which each patient is assigned to a sequence of treatments, including at least two treatments (of which one "treatment" may be a standard treatment or a placebo). Nearly all crossover designs have "balance", which means that all subjects should receive the same number of treatments and that all subjects participate for the same number of periods (39). The advantages and disadvantages of crossover trial design are given Tables 13.

Table 13. Advantages and disadvantages of crossover studies

Advantages	Disadvantages
• all subjects serve as own controls and error variance is reduced thus reducing sample size needed • all subjects receive treatment (at least some of the time) • statistical tests assuming randomization can be used • blinding can be maintained	• all subjects receive placebo or alternative treatment at some point; • washout period lengthy or unknown; • cannot be used for treatments with permanent effects

Issues to be considered in clinical trials are described briefly below.

Randomization:
As for cohort studies, subjects in RCTs are enrolled on the basis of their exposure status. In RCTs, however, the investigators themselves allocate the exposure. Participants in the two arms of RCTs should differ only with respect treatment (intervention) being trailed. In well-designed clinical trials, patients are assigned to the different arms of the trial using a system similar to flipping a coin. Such a grouping of patients by chance is called "randomization". In RCTs, randomization, if done appropriately, minimizes the potential for both bias and cofounding to far greater extent than is possible in any other study design (40). If sample size is large enough, then randomization should ensure that comparability between the intervention and control groups is obtained.

Blinding (Masking):
In some trials, patients know which arm of the study they are in. In other trials, patients do not know which arm they are in; this is called a "masked" or "blind" study. In some trials, neither doctors nor patients know which arm the patients are in and these trials are called "double-masked" or "double-blind" studies. Masking helps to minimize information bias.

Maintenance of compliance: RCTs require co-operation and active participation of study subjects. Participants may deviate from the study protocol for a number of reasons, including forgetfulness, misunderstanding instructions, developing side-effects, or just getting fed-up taking drugs/intervention. The extent of non-compliance in a trial is a function of both the duration and the complexity of the study. It is vital to make every attempt to ensure compliance or at least to measure it in order to assess the influence that poor compliance may have on the results of the study. Monitoring compliance is an important activity. Compliance may be assessed based on self-reports, pill counts or laboratory tests for biochemical markers as found applicable and feasible.

Ascertainment of outcome: Uniform ascertainment of outcome for both the intervention and control groups is essential in order to ensure that the results obtained are not biased by collection of either more complete or more accurate information from either of the study groups. Outcome for all (at least for most) of the study participants involved in the trial should be ascertained. If proportion of loss to follow-up patients is small and/or evenly distributed among both intervention and control groups, the risk of bias will probably be small. If this is high (say of the order of 30-40% or differs substantially between intervention and control groups, then the internal validity of the study is threatened. In clinical trials that are not out of necessity double-blinded, one would need to guard against (or at least measure) possible observer and/or subject biases.

Analysis and interpretation of RCTs: The first step in the analysis of RCTs is to compare the baseline characteristics of the treatment and control groups. The groups should be fairly comparable in terms of age, gender and other baseline characteristics where randomization has been effective. We then analyze the results and compare the frequency of outcomes between the two groups. All outcomes indicated as primary and secondary in the protocol should be analyzed and reported. The analysis should be by intention to treat (ITT) analysis of results of all those who have been allocated the various groups, not just analyze those who have completed a course of treatment (41). Modified intention to treat (mIIT) (42) and per protocol analysis (PP) (43) may also be considered and used as appropriate. Method for frequency outcome calculation is given in Table 14.

Sub group analysis requires due caution as one can obtain spurious results since various characteristics may not be evenly distributed among the sub-groups. Such group analysis may be used to generate hypotheses for suture studies, but otherwise should only be performed if they have been planned before the start of the study.

Table 14: Calculation frequency of outcome in Randomized Controlled trails

RESULT		Outcome	
		Present	Absent
Exposure to maneuver or putative causal agent	Yes	a	b
	No	c	d

Direction of Study

Outcome Frequency	**a** a + b	**Compared with**	**c** c + d

Ethical considerations in RCTs: It is also crucial that individuals are not denied access to treatment known to be effective just for the sake of involving them in a clinical trial. An end points committee may need to be established and assess results at various times in the course of the study. This will ensure that subjects are not harmed by either taking an unsafe treatment or being denied a useful treatment.

III. METHODS USED TO REVIEW SCIENTIFIC EVIDENCE:

SYSTEMATIC REVIEWS:

These studies review results from published individual studies. An extensive literature search is conducted to identify studies with sound methodology. Initially a wide net is cast to avoid publication bias but quality of the evidence reported is secured through stringent predetermined criteria to base analysis on those that met minimum standards. A systematic review cannot eliminate bias present in the original studies. Systematic reviews are given a lot of weight, and are at the top of the hierarchy of evidence because they look at a whole range of studies and draw conclusions from the findings (44).

META-ANALYSIS:

Meta-analysis is a statistical technique for combining the results of different randomized trials to see if the overall effect is significant (45). The analysis thoroughly examines a number of valid studies on a topic and mathematically combines the results using accepted statistical methodology to report the results as if it were one large study. It is important the researchers to include only numbers that are appropriate to combine. In addition, sometimes a meta-analysis is not advisable, for instance, when there are too many differences between the studies, there aren't enough studies, or the studies are missing a lot of data.

SUMMARY

This teaching article describes basic methods in clinical epidemiological studies and highlights common pitfalls and errors to be avoided. A well-articulated research question or hypothesis and use of an appropriate research method are critically important. The research design is then structured using an epidemiological method based on statistical principles. Data may be collected using a quantitative, qualitative or a mix of the two methods. Inferences are made on the relationships or associations of various 'exposures' to 'outcomes'. Descriptive and inferential statistics appropriate to the study strategy should be employed. The inferences made based on the analysis and the results and their accuracy depend on the accuracy of the data collected, the representativeness of the study sample, the power of the study and the accuracy of the statistical methods used. One should guard against bias that may be introduced at various stages of the study and the effect of confounding factors that could distort the results and conclusions of the study. Ethical considerations must be adequately addressed in any research undertaking. One should guard against methodological pitfalls or errors that would compromise the internal or external validity of epidemiological studies.

REFERENCES

1. Lulseged S. Editorial: Evidence to Inform Policy and Action: Going with the Flow. Ethiop Med J 2004;42(3).
2. Masic I, Miokovic M, and Muhamedagic . Evidence Based Medicine – New Approaches and Challenges**.** Acta Inform Med. 2008; 16(4): 219–225. Accessed on February 13 at doi: 10.5455/aim.2008.16.219-225 PMCID: PMC3789163
3. Godlee F. Milestones on the long road to knowledge. BMJ. 2007;334(Suppl 1):s2–3
4. Sackett DL, Rosenberg WM, Gray JA. Evidence based medicine: What it is and what it isn't. BMJ. 1996;312:71–2.
5. Fletcher R, Fletcher S. The Scientific Basis Clinical Medicine. In Fletcher R, Fletcher S (eds). Clinical Epidemiology: The Essentials, 4th ed. Lippincott: Williams and Wilkins: Philadelphia, PA 2005, pp2-3.
6. World Health Organization. Health Research Methodology: A Guide for Training in Health Research, 2nd ed. WHO Regional Office for the Western Pacific, 2001, pp9-10.
7. World Health Organization. The WHO Strategy on Research for Health. WHO, 2012, pp17-8.
8. Rothman J. Causal Inference in Epidemiology. Modern Epidemiology. Little, Brown and Company: Boston, MA 1986. pp7-21
9. Gabriel D. Inductive and deductive approaches to research. Accessed on February 13 at http:// deborahgabriel.com/2013/03/17/
10. World Health Organization. Health Research Methodology: A Guide for Training in Health Research, 2nd ed. WHO Regional Office for the Western Pacific, 2001, pp9-10.
11. Suresh K, Thomas S, SureshDepartment of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Trivandrum, India G. Design, data analysis and sampling techniques for clinical research. Ann Indian Acad Neurol. 2011 Oct-Dec; 14(4): 287–90).
12. World Health Organization. The WHO Strategy on Research for Health. WHO, 2012, pp17-8.
13. Write Content Solutions. Data Triangulation: How the Triangulation of Data Strengthens Your Research. Accessed on February 13 athttp://www.write.com/writing-guides/research-writing/research-process/data-triangulation-how-the-

triangulation-of-data-strengthens-your-research.
14. Ellis P. 2010. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge University Press. p. 52
15. Smith, GD; Ebrahim, S (2002) Data dredging, bias, or confounding. BMJ 325;7378:437-8. 5833
16. Hennekens CH, Buring JE. Epidemiology in Medicine. Little, Brown, and Company; Boston: 1987.
17. Sacket D. Bias in analytical research. J Chron Dis 1979;32:51-63
18. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Health C. 2016; 9: 211–7.
19. Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001;22:189-212.
20. Vanderweele T. On the distinction between interaction and effect modification. Epidemiology 2009; 20:863-71
21. Heffner CL. Chapter 7.3 Test Validity and Reliability. August 21, 2014-11-22T02:53:24+00:00
22. Suresh K, Thomas SV, Suresh G. Design, data analysis and sampling techniques for clinical research. Ann Indian Acad Neurol. 2011 Oct-Dec; 14(4): 287–90
23. Kimberlin CL,Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008 Dec 1;65(23):2276-84.
24. Pagano M. Gauvraeau. Data Presentation (Chapter 2). Principles of Biostatistics. Wadsworh Publishing Company; Belmont, CA: 1993. Pp7-30
25. Creative Research Systems. Significance in Statistics & Surveys. Accessed on February 13 athttp:// www.surveysystem.com/signif.htm
26. Institute of Work and Health. What researchers mean by statistical significance. Accessed on February 12, 2017 at https://www.iwh.on.ca/wrmb/statistical-significance
27. World Medical Association (WMA) Declaration of Helsinki – Ethical Principles of for Medical Research Involving Human Subjects. Adopted by the 18th General Assembly, Helsinki, Finland, June 1964 and Amended by WMA Assembly, Fortaleza, Brazil, October 2013
28. National Research Ethics Review Guideline. FDRE Federal Ministry of Science and Technology 5th Ed. 2014
29. Council for International Organizations of Medical Sciences (CIOMS) 2002. International Ethical Guidelines for Biomedical Research Involving Human Subject. Accessed on December 28 at http://www.cioms.ch/ frame_guidelines_nov_2002.htm (1 de 64)08/03/2007 9:10:05
30. World Health Organization 2011, Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants. Accessed on December 28 at http://apps.who.int/iris/ bitstream/10665/44783/1/9789241502948_eng.pdf?ua=1&ua=1
31. Crimes D, Schultz, K. (2002) "Descriptive Studies: what they can and cannot do". The Lancet, 359,145-9.
32. Health Knowledge: Education, CPD and Revalidation from PHAST. Accessd on February 123, 2017 at https:// www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/cs-as-is/cross-sectional-studies
33. Whitaker H. Case series studies: rationale, design and analysis. Open University, UK. h.j.whitaker@open.ac.uk. Access on February 13, 2017 at https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/research_ presentations/ Whitaker_case_seri.
34. Schulz K, Grimes D. Case-control studies: research in reverse. The Lancet 2002;359(9304): 431–4
35. Pearce N. Analysis of matched case-control studies. BMJ 2016;352:i969 doi: https://doi.org/10.1136/bmj.i969 (Published 25 February 2016) Cite this as:
36. Grimes Ds, Schulz K. Cohort studies: marching towards outcomes. The Lancet 2002;359(9303):341–5.
37. Ernster VL. Nested case-control studies. Prev Med. 1994 Sep;23(5):587-90
38. MacGil M. What is a randomized controlled trial in medical research? Updated on January 19, 2016. Accessed on February 13, 2017 at http://www.medicalnewstoday.com/articles/280574.php
39. Johnson D Crossover experiments. WIREs Comp Stat 2010;2: 620-5.
40. Kunz R, Vist G, Oxman AD. Randomization to protect against selection bias in healthcare trials. Cochrane Database Syst Rev. 2007 Apr 18;(2):MR000012.
41. Hollis S, Campbell F.What is meant by intention to treat analysis? Survey of published randomized controlled trials. BMJ. 1999 319(7211): 670–4.
42. Montedori M, Bonacini M, Casazza G, et al. Modified versus standard intention-to-treat reporting: Are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. Trials 2011;12:58. doi: 10.1186/1745-6215-12-58
43. Shah PB. Intention-to-treat and per-protocol analysis. CMAJ. 2011 183(6): 696. doi: 10.1503/cmaj.111-2033
44. Khan K, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review . R Soc Med. 2003 Mar; 96 (3):118–121.PMCID: PMC539417.
45. Haidich AB. Meta-analysis in medical research. Hippokratia. 2010 Dec; 14(Suppl 1): 29–37