Chapter-29 Introductory Research Methodology

Chapter · January 2016

DOI: 10.5005/jp/books/12773_30

3 authors:

Narendra K. Arora
Independent Researcher
251 PUBLICATIONS **3,369** CITATIONS
SEE PROFILE

Rakesh Pillai
INCLEN Trust
21 PUBLICATIONS **87** CITATIONS
SEE PROFILE

Sanjiv Kumar
Indian Academy of Public Health
55 PUBLICATIONS **562** CITATIONS
SEE PROFILE

Social Pediatrics: Introductory Research Methodology

Narendra K Arora, Rakesh N Pillai
The INCLEN Trust International, F-1/5, Phase 1, Okhla, New Delhi, 110020
Email: nkarora@inclentrust.org

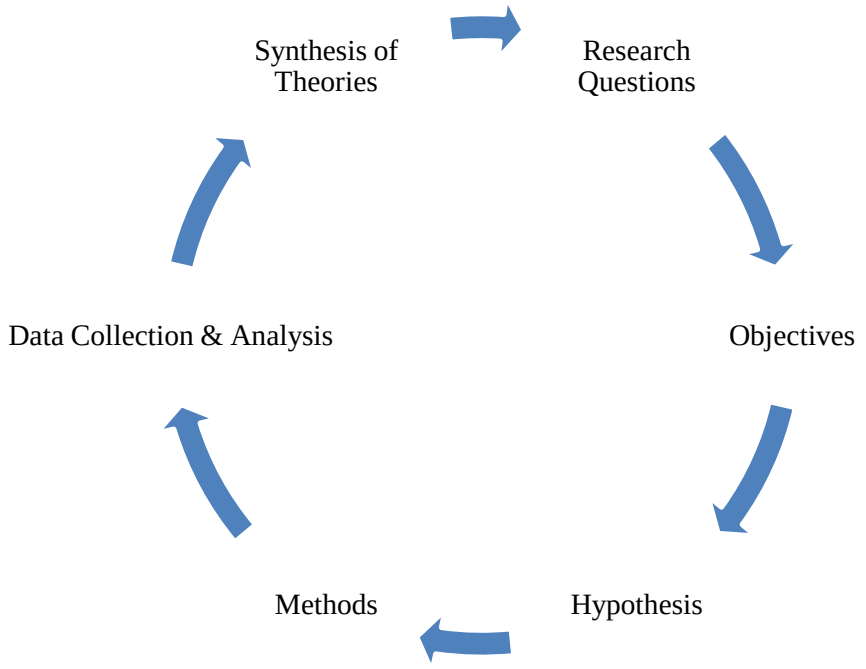
What is the rationale for doing medical research?

Medical research is conducted for creating new knowledge architectures or to generate fresh interpretation of known facts and principles focusing on high impact health outcomes. Research is formally defined as *'a quest for knowledge through diligent search or investigation or experimentation aimed at the discovery, development and interpretation of new knowledge'*.

Researchers have to follow systematic procedures and techniques in carrying out the search, investigation or experimentation to improve on the existing knowledge and/or obtaining new knowledge, broadly referred in as research methodology. Research is also the cornerstone for informed and effective decision-makings, for diagnosis and treatments by the health care providers and for designing and refining policies and programs by the policy makers and program managers. Thus researchers should focus to influence stakeholders across health and non-health sectors so that their actions should converge to enhance human, health and economic development of society.

Advancement in technologies has impacted research outlooks. For example recently public health data are managed through advanced informatics infrastructures for creating a cycle of real time learning's that enable individuals, community, organizations and networks, government, and corporations. Population health records and big data from multiple sources are widely used to identify early signals of key public health issues. Effective communication to peers, health practitioners and community at large is also the responsibility of medical researchers. Therefore all researchers should understand the basic philosophy of medical research follows a iterative nature.

Fig-1: Iterative nature of research



Different type of research

Based on the underlying philosophies research could be categorized as following:

1. **Empirical and theoretical research**

Empirical research that involving documentation of observation, experience and analyzes them to arrive at specific conclusions. Theoretical research that involves development of a theory and abstracts based on the available evidences or knowledge and their variations. Most of the public health research will fall under the empirical research category. Empirical and theoretical research complement each other in developing common understanding of a phenomena, in predicting future events, and in the prevention of events harmful to the general welfare of the population of interest. Empirical research assists in confirming or refuting the theories.

2. **Basic and applied research**

As stated earlier, research can also be categorized into basic (pure) and applied research. Basic research is usually a systematic enquiry directed towards generating and/or improving knowledge of the fundamental aspects of a phenomenon and their observable facts without specific application towards processes or products in mind. Applied research aims to find solutions for a problem faced by a society or health system or industry. This is also called as action research. In short, basic research discovers principles and theories while applied research discovers ways of applying them to solve specific problems. A suitable balance between these two categories of research is needed for bringing a wholesome benefit to the society and system.

3. **Descriptive and Explanatory Research**

Descriptive research often aims to describe a situation(s), event(s) or social system(s). It describes the state of affairs as it exists. Surveys and fact-finding enquiries of different type are part of descriptive research. In descriptive research, the researchers have no control over variables and report only what has happened or is happening. Explanatory research, aims to establish the cause and effect relationship and also the nature of the relationships. The researcher analyzes the facts or information for assessing and/or establishing these relationships.

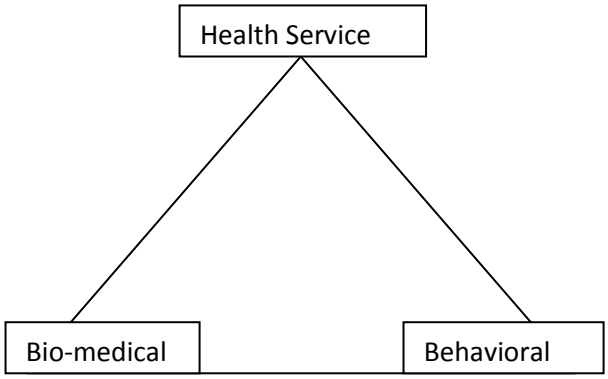
4. **Quantitative and Qualitative Research**

Quantitative research is the systematic and scientific investigation of quantitative properties and phenomena and their relationship through numerical data. The objective of quantitative research is to develop and employ mathematical models, theories and hypotheses. A fundamental principle in quantitative research is that underlying correlations does not imply causation. This principle follows from the fact that it is always possible for a spurious relationship to exist for variables between which covariance is found in some degree.

Qualitative research aims to gain a deeper understanding of a specific topic or event or society or organisation. This methodology studies things in their natural setting and attempt to make sense or to interpret phenomena in terms of the meanings people bring to them. Qualitative research can provide rich information about any social processes in a specific setting. It also complements the findings from quantitative research and helps to understand some of the outcomes and association.

5. Health research triangle

Another way of categorizing health research; be it empirical or theoretical, basic or applied is to describe it under three interlinked operational categories of biomedical, health service and behavioral research the so-called health research triangle.



Biomedical research deals primarily with basic information involving processes at the cellular and tissue levels while health service research deals with issues in the human environment and population at large which promote changes at the cellular level and behavioral research deals with the interaction of humans and population with the environment in a manner reflecting the beliefs, attitudes and practices of the individuals in the society.

Good practices in medical research [(i)]

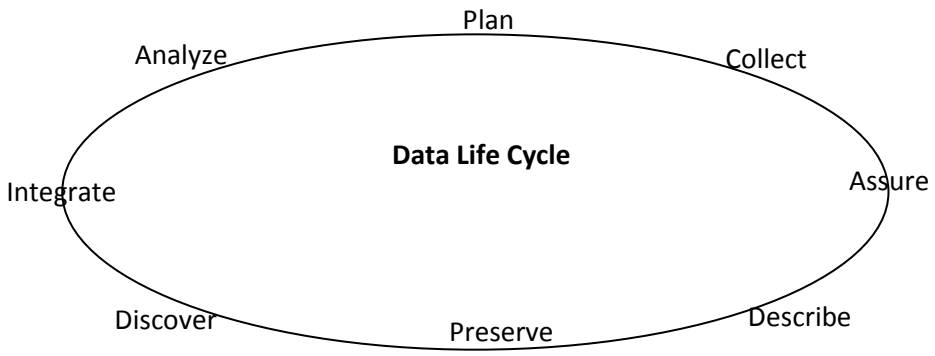
Most of the countries have set country specific regulations for conducting research; however the following 12 principles serves as the foundation for all good practices in research. All researchers should follow:

1. Principles of essentiality
2. Principles of voluntariness, informed consent and community agreement
3. Principles of non-exploitation
4. Principles of privacy and confidentiality
5. Principles of precaution and risk minimization
6. Principles of professional competence
7. Principles of accountability and transparency
8. Principles of maximization of public interest and of distributive justice
9. Principles of institutional arrangements
10. Principles of public domain
11. Principles of totality of responsibility
12. Principles of compliance

Data and its effective use is the cornerstone for all good practices in research. All research should have a well coordinated planned resource management protocol throughout its data life cycle. Often multiple data sources have to be linked to derive meaningful conclusions of outcomes. Researchers need to be familiar with few definitions in data management. Facilities used to store biological data are called as bio-banks. Digital data derived from any bio-medical research is defined as bio-data.

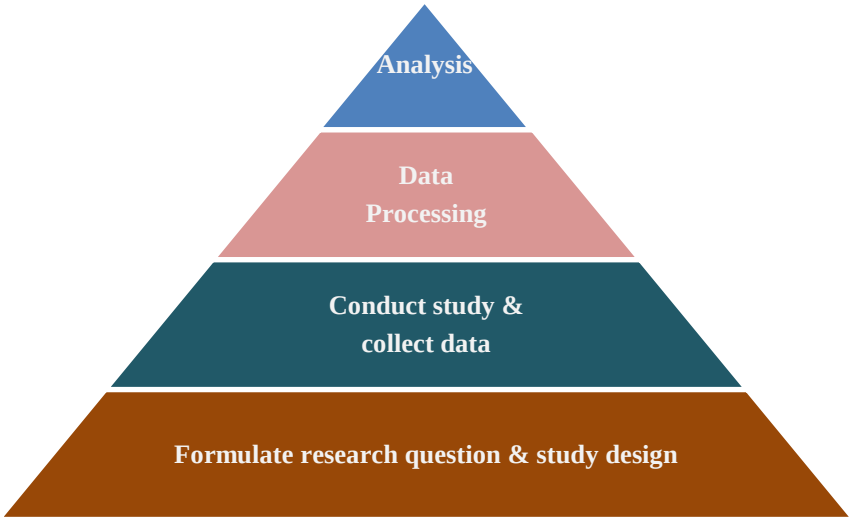
Any biological material collected from a study participant is called bio-specimen. Bio-repositories are facilities used to store human bio-specimens.

Components of a data life cycle are;



Under good practices, researchers have to comply with all legal, ethical and social guidelines as well as efficiently manage practical issues such as data collection, storage, retrieval and sharing of information. Implementing complex informed consents is also the responsibility of a researcher. Any plan for secondary analysis, executing material transfer agreements (if any), dealing with intellectual property issues, obtaining IRB approvals and processes to ensure compliance with human rights and bio-safety regulations are all part of researcher responsibility.

Fig 2: Theoretical work flow of research



Defining the research problem

Properly defining a research problem could be perceived as the first challenge to a researcher. Creating a document (e.g: a protocol or a proposal) will always help researcher to systematically do this. Researcher should start with writing the background of the problem and move on to the rationality in operational terms. For this, two steps are essential; (i) understanding the problem thoroughly and (ii) rephrasing it into meaningful terms from an operational point of view.

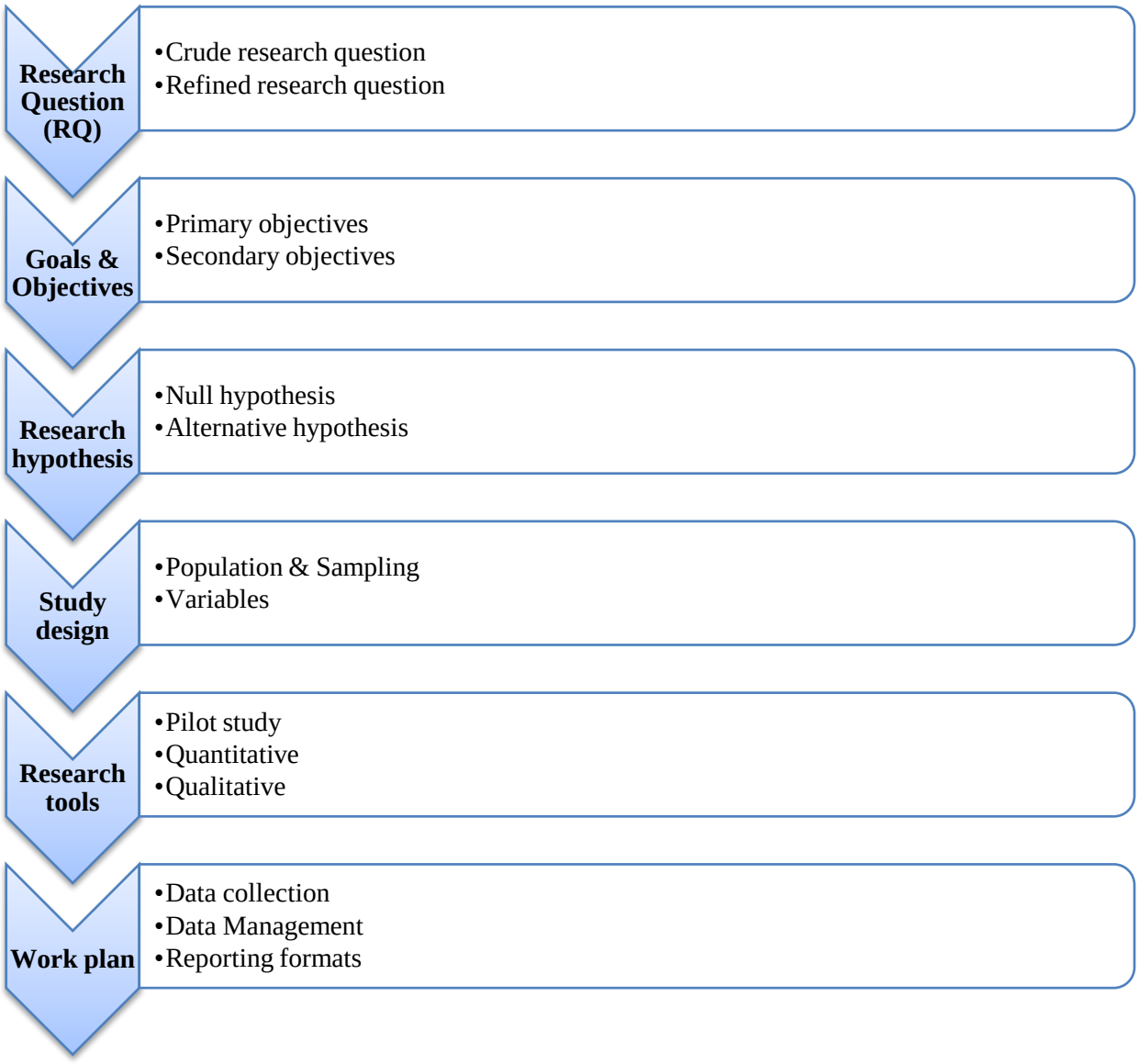
Therefore all research should start with a clear statement of the research problem which should:

- be based on observation and experimentation on theories (empirical)
- build on existing data using both positive and negative findings and collect new data and organize in such a way that they answers the original research question (systematic)
- apply critical analysis and interpretation based on real-time observations and statistical significance
- maintain neutrality, un-biased and logical thinking
- consider ways to control information to highlight the difference within and between groups
- consider qualitative enquiries to generate relevant hypothesis
- also consider quantitative investigations to test hypothesis

Research Question

After defining a research problem the logical step is to develop a research design. First step in this is to frame a research question which is the logical guide for the research. It is important that research questions should address the following four elements in the acronym i.e., **PICO**: Patients, Intervention (for intervention studies only), Comparison group, and Outcomes. Some also suggest for adding T for Time (thus making it PICOT).

Fig-3: illustrates key points to be considered while preparing a research question

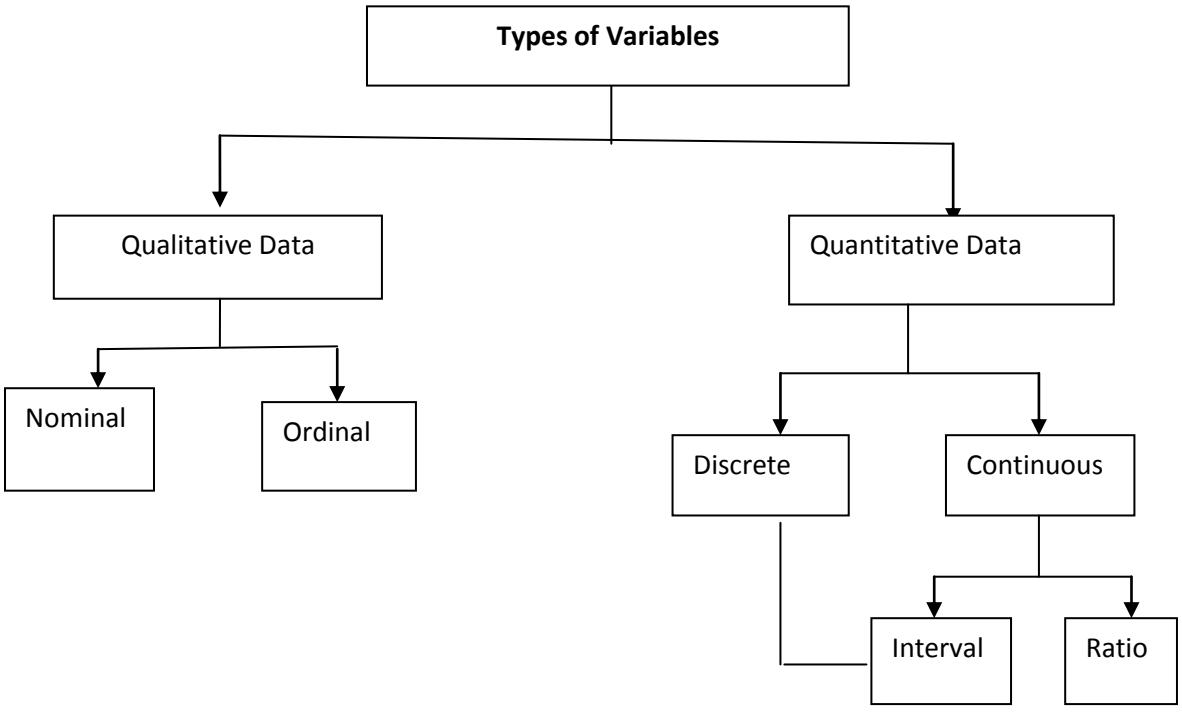


To understand research question in detail let us get used to the following terminologies:

a) **Observation**: Each person or a thing for which the data is collected is called an observation. In bio-medical research these are often peoples or subjects.
b) **Constants:** If a characteristic is same for all observation it is called a constant.
c) **Variable:** If a characteristic of an observation changes it is called a variable. In broad sense a variable is an entity which could take different values for different people. In medical research, variables are things that we measure, control or manipulate. For example, height, weight, income etc.
    - **Independent variable** (also called *predictor variable*): These are variables which are manipulated or controlled or changed during the course of the research
    - **Dependent variable** (also called *outcome variables*): These are variables which changes on account of independent variables

      For example: In a research question, whether vitamin supplementation could extend life expectancy of people, then independent variable is the amount of vitamin given to the person and the life span of the person is the dependent variable.
    - **Extraneous variable:** Independent variable that are not related to the purpose of the study but may affect the dependent variable is termed as extraneous variable
    - **Moderator:** This is a third variable when introduced into the analysis will alter or produce contingent effect on the cause-effect relationship between independent and dependent variables
d) **Dataset**: A dataset is the collection of several pieces of information called variables (often arranged in columns).

Variables could be of different types based on the datasets as mentioned below:



Hypothesis

Next step in designing the research protocol is to formulate a relevant hypothesis. A hypothesis is defined as *'scientific reasoning for a phenomenon so as to design the study to prove or disprove it'*, which is formulated in one of two possible directions: induction or deduction.

Inductive methods are those when data (observations) is initially collected and analyzed to reach out to a hypothesis (theory) while deductive methods emphasizes that theory comes first. Inductive methods are more common in qualitative research while deductive methods are common in experimental and analytical research. However in broad sense both of these philosophies co-exist in a clinical or a public health research.

A hypothesis should be considered as the fundamental tool in a research which goes hand in hand with the selection of a research problem. Empirical researches are perhaps the most common source of generating hypothesis; which is the product of experience and rational thinking about the subject. Hypothesis is more important in experimental research where researchers often make prediction about the outcome of his/her experiment. On the other hand in a descriptive research, researchers seek facts and figures for a particular situation which do not require a prominent hypothesis. Having said so, most of the descriptive studies not only involve the analysis of a situation but also make generalizations from an observed fact, condition or a behavior. Therefore it is recommended to use a hypothesis wherever possible. Good hypothesis comes from experience and creativity which require thorough review of literature and should satisfy following criteria:

1. must be clear and precise
2. must be specific and use simple terms
3. should be capable of testing in a specific time
4. should state relationship between variables
5. should be consistent with known facts

Good hypothesis will help researchers to select: appropriate sample, design a data collection protocol, statistical test and establish relationship between variables. There are different types of hypotheses named according to its usage and it is important to differentiate these hypotheses.

a) **Research hypothesis** – It is designed at the beginning of a research summarizing the broad question and the objective of research is to test a statement using scientific methods.
b) **Statistical hypothesis** – This is often used in relation to the population under study and has to be translated into testable forms in the course of study
c) **Null hypothesis** – This is a statement to prove or disprove a hypothesis using statistical test, thus this is also called under statistical hypothesis. Researchers state that 'there is no significant difference' between the variables. Each variables tested would automatically imply an inherent statists in it.
d) **Directional hypothesis** – This stipulate the direction of expected differences in the groups
e) **Alternative/Declarative hypothesis** – This indicates a situation when null hypothesis is not true. It is called declarative because researcher makes a positive statement about the outcome. Null and alternative hypothesis is mostly important in calculating the power of research and the sample size.

Hypothesis testing is a method for testing a claim or hypothesis about a parameter in a population using data measured from a sample. This is often used in quantitative research as a decision making process from the sample data. At some occasion it is also called confirmatory data analysis. Hypothesis test could be done experimentally and non-experimentally.

The principles of hypothesis testing could be summarized in four steps as;

Step 1: Stating a hypothesis (null as well as alternative)
Step 2: Setting a criterion for decision
Step 3: Do a test statistic
Step 4: Decision making

This could be explained with an example. A researcher wanted to test a claim found in a literature ***'obese children watch TV for 3 hours per week'***. Overall objective of the researcher is to prove or failing to prove the likelihood of the population estimate (3 hours per week of watching TV by obese children). For this, researcher will start with a null hypothesis which is presumed as true at the beginning of the study. In the above example null hypothesis would be ***'obese children TV watching time per week is not equal to 3 hours'***. Here the average time of TV watching by obese children could be greater 3 or less than 3 however the direction of the results are not known. Remember that the researchers assume that null hypothesis is true and the objective of researcher is to conduct a study and apply **'test statistic'** as an evidence to show that null hypothesis is not accepted. That means the alternative hypothesis (which is the claim opposite to null hypothesis) may be true.

Next step is to set decision making criterion that is called as '**level of significance**'. It is the criterion selected to prove or failing to prove the null hypothesis. The question is how much difference (variance or discrepancy) from the mean (population mean i.e., 3 hours per week) is acceptable in our study (sample mean) so that the null hypothesis we are testing could not be proved. In medical research it is often set as 5% significance level. This means that when the probability of obtaining a sample means (not equal to 3 hours as in null hypothesis) is less than 5% then we fail to prove the null hypothesis. Alternative hypothesis is important when the direction of difference between sample mean and population mean is not known.

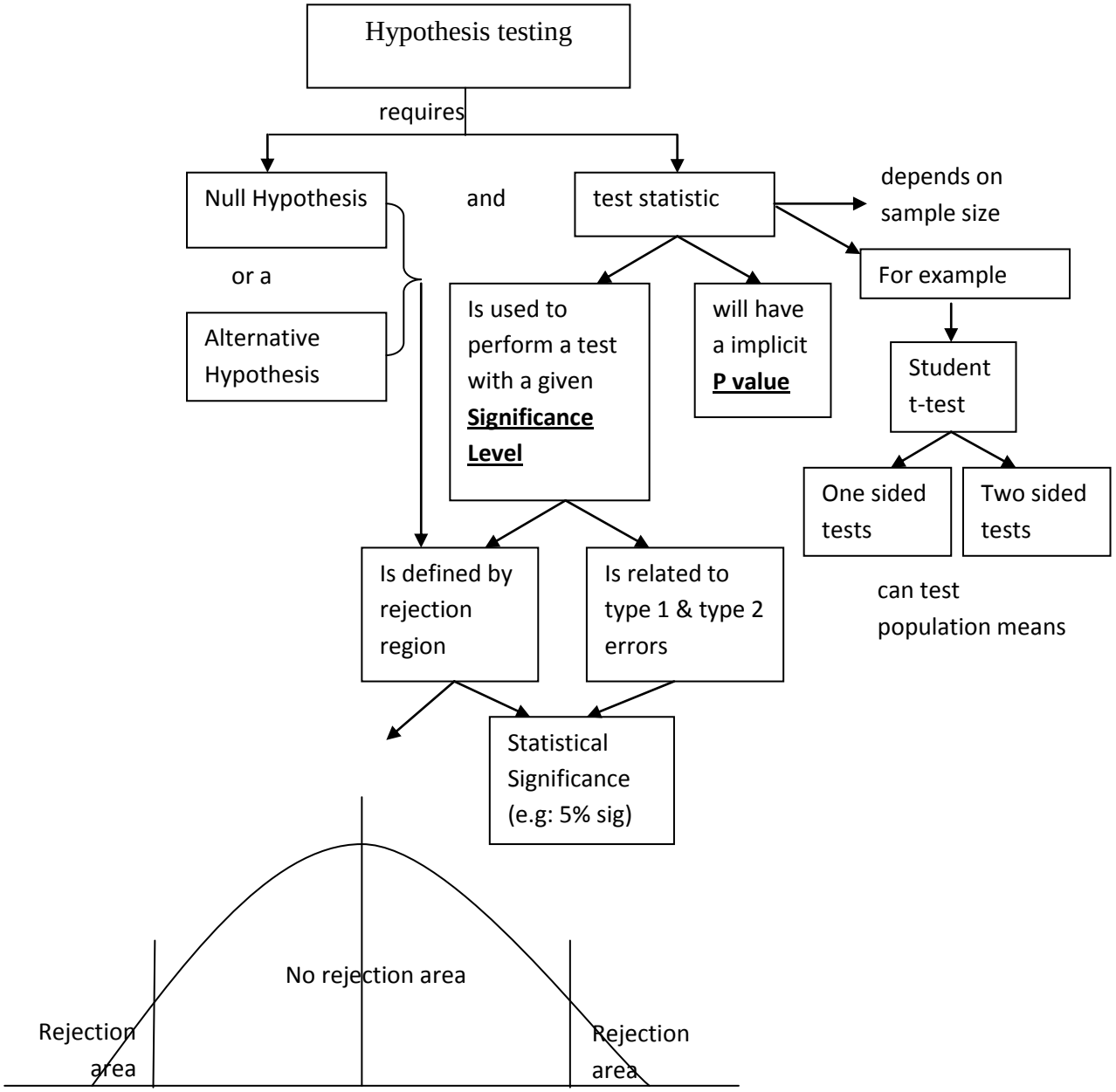
A '**test statistic'** is the mathematical formula that allows a researcher to determine the likelihood of obtaining sample outcome if null hypothesis were true. It will help to understand the likelihood of standard mean equals population mean. It tells us how many standard deviations a sample mean varies from a population mean. Larger the value of test statistic, further the distance or number of standard deviation of sample mean is from population mean in null hypothesis. Therefore it is important to select relevant test statistic which is often based on the 'measurement scale of data'. Few considerations in selecting the measurement scale and appropriate test statistics are summarized in the table below. After applying a test statistic, if the probability of obtaining sample mean is less than 5% when the null hypothesis is true then the decision is to reject the null hypothesis while if probability of obtaining sample mean is greater than 5% then the decision is to fail to reject the null hypothesis. This is decided by p values or significance tests.

Table 1: Measurement scale of data and common type of test statistics

	Specification	**Type of data**	**Examples**	**Example of test statistic used**
Nominal Scale	Will be names	Unordered categories	Type of therapies	Chi-square
Dichotomous	Type of nominal scale which have only two categories	Binary data	Gender: male or female	Chi-square or Fishers exact test, Mc-Nemar test
Ordinal Scale	Will have names with a specific order	Ordered categories	Quality of life	Wilcoxon Mann whitney test
Interval Scale	Will have names in an order with equal intervals and an arbitrary zero point	Continuous with arbitrary zero point	Body temperature	T test (if related samples paired t test)
Ratio Scale	Will have names in an order with equal intervals and an absolute zero point	Continuous with absolute zero point	Adherence : 0% means there is no adherence	Chi-Square test
Continuous		Interval & ratio scale		Z test, T test

The probability of obtaining a sample mean, given that the value stated in the null hypothesis is true is called **'p value'**. The p value is the probability which varies between 0 and 1. P value for obtaining a sample outcome is compared to the level of significance. When 'p values' is less than 5% ($p<0.05$), researcher reach significance, the decision is to reject null hypothesis. 'Significance' or 'statistical significance' is the decision made based on the value in null hypothesis. When null hypothesis is rejected we reach significance and if we fail to reach significance the null hypothesis is failed to be rejected.

Fig-4: concept of hypothesis testing



Type 1 Error (α)**:** It is the probability of rejecting a null hypothesis that is actually true. In real life situation it is like blaming a person guilty when he is not actually guilt. Level of significance is actually stated to control this error and the level we set for this is called **alpha level** (α). It is defined as the largest probability of committing a type 1 error that is allowed in the particular study and still decides to reject null hypothesis.

Type 2 Error (β)**:** It is the probability of failing to reject a false null hypothesis.

Power (1-β): This is the correct decision taken to reject a null hypothesis when it is really not true and is called as power of the decision making process. It is called as power because this is the decision we aimed. To be more specific it is the probability that a randomly selected sample will show the null hypothesis as false when it indeed false. This is explained in a 2X2 table below:

	Failing to reject the null hypothesis	**Rejecting the null hypothesis**
True	Correct (1-α)	Type 1 Error (α)
False	Type 2 Error (β)	Correct (1-β)

Measures of Disease Frequency

A pre-requisite for medical or public health research is the ability to quantify the occurrence of an event of a disease. This is achieved through investigating the extent of a problem, identifying patterns and trends in disease occurrence, identifying the cause of disease relationship and evaluating the effectiveness of prevention and treatment activities. Therefore it is important to understand the principle behind few mathematical terms commonly used in research.

1. **Ratio:** It is the comparison of two numbers. It is obtained by simply dividing one quantity by another without implying any relationship between numerator and denominator **(a/b)**.
   For example: number of still births per thousand live births.
2. **Proportion:** It is the part or amount considered in relation to a whole. Proportion is a type of ratio in which those included in the numerator is also included in the denominator **(a/(a+b))**.
3. **Rate:** This is a measure, quantity or frequency, typically one measured against another during a certain time period. Thus rate is a ratio in which a distinct relationship between the numerator and denominator exist. Most essentially ratios have a measure of time as an intrinsic part of the denominator.

Rates and proportion is always a ratio however a rate may or may not be a proportion.

Risk: Risk is a rate expressed in relative frequency of an event per unit time.

Few rates are commonly used in public health or epidemiological research which is as mentioned below:

$$\textbf{Infant mortality rate (IMR)} = \frac{\text{total \# of infant deaths } (< 12\ months\ old) \text{ in a year}}{\text{\# of live births in that year}} \text{X } 1000$$

IMR is a ratio however this may not be a proportion because the numerator is not necessarily part of the denominator (some infants might have born in the previous year).

$$\textbf{Neonatal mortality rate (NMR)} = \frac{\text{number of child in first 28 days in a given year}}{\text{\# of live births in that given year}} \text{X } 1000$$

$$\textbf{Fertility rate} = \frac{\text{total \# of live births in a year}}{\text{live births in women in reproductive age in that year}} \text{X } 1000$$

Fertility rate is both a ratio and a proportion.

$$\textbf{Crude death rate (yearly)} = \frac{\text{total \# of deaths in a calender year}}{\text{midyear population}}$$

$$\textbf{Age specific death rate (for under 5)} = \frac{\text{total \# of deaths in} < 5\ years\ in\ a \text{ calender year}}{\text{midyear population of age} < 5\ years}$$

At times ratios are expressed per 100, 1000 or 100,000 persons depending on the strength of observation.

$$\text{Death rate for age } 1-4 = \frac{\text{total \# of deaths aged } 1-4 \text{in a calender year}}{\text{total \# of deaths in calender year}} \text{ X } 100$$

$$\textbf{Incidence rate} = \frac{\text{\# of new cases of specific disease in a calender year}}{\text{total mid year population}} \text{ X } 100$$

$$\textbf{Prevalence rate (point)} = \frac{\text{\# of cases [old \& new]of specific disease at time 't'}}{\text{total population at time 't'}} \text{ X } 100$$

$$\textbf{Period prevalence} = \frac{\text{\# of cases diagnosed with a specific disease in a time period}}{\text{total population in that time period}} \text{ X } 100$$

Rate ratios / relative rate: It is the ratio of two events in person-time, for example incidence. In other words it is the measure of association of time varying exposures with occurrence of an event given in person times (often used in the cohort studies). It is the direct measure of risk (probability) that health person will develop a disease during a specific period of time. Hence it is the analogue of ***risk ratio*** for person time measures and is the basic tool in causality assessment (etiology). Often rate ratios are expressed as ***relative risk***.

Relative risk is formally defined as the ratio of diseases in people who are exposed to the presumed cause to the incidence among those who are not exposed. Incidence of a disease means absolute risk. For a cohort study with total data count in the denominator, the relative risk is calculated as the ratio of cumulative incidence among those exposed compared with those not exposed.

A relative risk of 1.0 indicates that incidence rate of disease in the exposed and unexposed group are identical and thus there is no association observed between the exposure and the disease in the data. A value greater than 1.0 indicates a positive association, or an increased risk among those exposed to a factor (indicating causal relationship). A value less than 1 indicates protective effect in exposed as compared to non-exposed. For example, a relative risk of 1.8 to detect zinc deficiency among malnourished children would mean that zinc deficiency is 1.8 times or 80% more likely if the child is malnourished than well nourished. It is also important that incidence rate should not be compared to cumulative incidence. Cumulative incidence is the proportion of people who convert, during a specified period of time, from non disease to diseased. It is difficult to interpret one of each type of incidence measure.

Table 2: 2 X 2 tables for risk estimation

	Disease	**Non diseased**	
Exposed	A	B	A+B
Not exposed	C	D	C+D
	A+C	B+D	A+B+C+D

Incidence among exposed ($Risk_{exp}$) = $\frac{a}{a+b}$

Incidence among non exposed ($Risk_{unexp}$) = $\frac{c}{c+d}$

Relative risk = $\frac{\frac{a}{a+b}}{\frac{c}{c+d}}$

Odds Ratio (OR)

'Odds' of an event is defined as the probability of an event to occur compared to the probability that event will not occur. **Odds ratio** is the odds of that event to occur. In other words it is the odds that an outcome to a particular exposure will occur compared to the odds of that outcome occurring in the absence of that exposure. It is a measure of association between an exposure and an outcome variable. This is also a common method for displaying risk. 'Odds' of a person developing a disease is his/her chance of acquiring the disease divided by the chance of not acquiring that disease; in either exposed or non exposed groups. OR is commonly used in case-control studies however they could also be used in cross-sectional and cohort study designs with some assumptions.

In statistics, OR are used to compare the relative odds of the occurrence of an outcome of interest to its variable of interest. The odds ratio can also be used to determine whether a particular exposure is a risk factor for a particular outcome and to compare the magnitude of various risk factors for that outcome.

Therefore,

Odds of an exposed person acquiring disease = $\frac{\frac{a}{a+b}}{\frac{b}{a+b}} = \frac{a}{b}$

Odds of a non exposed person acquiring disease = $\frac{\frac{c}{c+d}}{\frac{d}{c+d}} = \frac{c}{d}$

Thus odds ratio = $\frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$

In case if 'a' and 'c' are rare events (<1% incidence) then we can safely assume that relative risk equals odds ratio. An odds ratio of 1 indicates that the condition or event under study is equally likely in both groups. An odds ratio greater than 1 indicates that the condition or event is more likely in the first group and an odds ratio less than 1 indicates that the condition or event is less likely in the first group.

For case control study designs where most often the disease incidence is considerably lower than 1% odds ratio is a good approximation to relative risk. That is OR provides a valid estimation of the relative risk in conditions that prevail in most of the case control studies including the cases of disease that are newly diagnosed, provided prevalent cases are not included and that the selection of cases and controls are not based on exposure status. This could be further explained with an example of diarrheal deaths in infants with breast feeding as exposure [Source: example from internet].

A study has analyzed the information of 170 infants who died due to diarrhoea and 340 infants who survived to the 2nd year for their history of breastfeeding. The results are as mentioned below.

	Diarrhea	**Controls**	**Total**
No Breastfeeding	120	136	256
Breastfeeding	50	204	254
Total	170	340	510

Odds ratio of infant death to breast feeding = (120 X 204 / 136 X 50) = 3.6

Interpretation: It could be explained as odds of not been breastfed are 3.6 times among infants who died due to diarrhea in comparison to those survived into the 2nd year.

Often this type of data is collected in case control studies. In the above example the exposure-odds ratio is equal to the disease-odds ratio. Since the disease is rare and the exposure-odds ratio is almost equal to the relative risk this information could be related to relative risk in cohort studies.

In the above example stratifications can be done based on age, sex, socio-economic status etc. When stratification is done then **interaction or effect modification occurs**. Association between risk factors and outcomes could be studied using stratifications.

For example, if the cases were stratified according to the age of infants, as less than 1month and above 1 month of age then;

Strata 1: Infants < 1 month of age

	Diarrhea	Controls	Total
No Breastfeeding	10	3	13
Breastfeeding	7	68	75
Total	17	71	88

The odds ratio = (10X68 / 3X7) = 32.4

Strata 2: Infants >= 1 month of age

	Diarrhea	Controls	Total
No Breastfeeding	110	133	243
Breastfeeding	43	136	179
Total	153	269	422

The odds ratio = (110X136 / 133X43) = 2.6

In the above example, odds of deaths due to diarrhea are much higher in neonates as compared to other infants (32.4 vs. 2.6).

Calculate incidence and prevalence rates

Age-specific number of cases of typhoid cases (culture-confirmed) detected from active surveillance over a 1-year period in urban area is given below.

Age of follow-up (years)	Total follow-up (years)	Culture-confirmed (cases)	Calculate Incidence..?
Under 5	1027	28	
0-1	166	0	
>1-2	202	5	
<2-3	213	11	
>3-4	225	5	
>4-5	221	7	
Over 5-19	273	32	
≥ 5-12	1579	22	
>12-19	1164	10	
Over 19-40	2684	3	
Total	6454	63	

Attributable risk

Attributable risk may be defined as the difference between the rates at which disease occurs in people who are exposed and in people who are not exposed.

$$AR = (Incidence\ rate\ in\ exposed\ population) - (Incidence\ rate\ in\ unexposed\ population\ )$$

The public health importance of attributable risk is that it is a measure of additional risk of the disease in the exposed individuals and conversely the reduction in the disease incidence which may be expected from removal of the exposure. The term attributable risk is unwarranted if there is no cause-effect relationship between exposure and disease. Dividing the attributable risk with the rate of occurrence in exposed groups gives the etiologic fraction or attributable risk percent.

$$\text{Attributable/etiologic fraction} = \frac{\text{I exp} - \text{I unexp}}{\text{I exp}}$$

It conveys a sense of how much of the disease in the exposed population can be prevented by blocking the effect of the exposure or eliminating the exposure.

Population attributable risk

Population attributable risk is derived exactly the same way as attributable risk and estimates the potential effect of changes in a society. For example, effect of elimination of cigarette smoking or reduction of cholesterol on prevalence of coronary artery disease etc. It is the rate of specific outcome in a population that can be attributed directly to the disease and is calculated by following equation:

$$Population\ AR = \\ rate\ of\ an\ outcome\ oin\ whole\ population - rate\ of\ outcome\ in\ non\ exposed\ population$$

Or

$$= attributable\ fraction\ X\ prevelance\ of\ risk\ factors\ for\ theoutcome\ of\ whole\ population$$

Population attributable risk is sensitive to the incidence rate or prevalence of disease in the target population. However it is not affected by changes in the baseline incidence of the disease. The magnitude of the relative risk depends on the magnitude of the baseline incidence rates. The same attributable risk in two populations can correspond to greatly different relative risks conversely the same relative risks for two populations could correspond to greatly differing attributable risk.

Establishing Causal Relationship

Establishing causal relationship is not from the direct result of a study. This is the most commonly misunderstood concept in research in the attempt to adding legitimacy to research. The investigator has to answer the question "is the exposure from the study really causes disease in the population at risk". The answer to this is not in statistics but it depends on the applied logic and common sense. The key principle of establishing cause and effect relationship is by proving that the effects seen in the study happened after the cause (defined as the temporal relationship). Researcher has to keep in mind that it is not possible to establish complete causality however every researcher have to strive to establish 100% proof of causality. Few headings have to be remembered while establishing cause-effect relationship as:

1. Strength of association
2. Consistency
3. Specificity
4. Temporality /time relationship
5. Dose-response relationship / biologic gradient
6. Biologic plausibility
7. Experiment
8. Analogy
9. Coherence of evidence

This could be better explained with an example. If we have to study the exposure to benzene and occurrence of leukemia in industrial workers then;

1. **Strength of association**: In addition to the study results, researchers have to look at the literature evidences to prove the strength of association. For example the relative risk of

leukemia in industrial workers, in literatures, ranges from 3.8 to 25. Is this a strong association? Then how strong is strong? This has to be decided by the investigators.

2. **Consistency**: Investigators need to prove that the relationship found in the study is more than a chance finding. One would like to give examples of the same relationship being demonstrated on many occasions, in different work situations, distinct geographic regions and ethnic groups and at different times. In the above example of benzene exposure and leukemia, most studies in relationship between benzene and leukemia show positive relationships, though some studies have failed to show a definite relationship. Moreover none of the studies have shown a protective effect of benzene. Researchers need to evaluate the studies for their methods.
3. **Specificity**: The researcher has to establish a cause-effect relationship also. For example, whether benzene produce only leukemia or if leukemia was only due to benzene. This condition was not met with even in the strongest causal relationship.
4. **Temporality**: As mentioned above, cause must precede an effect and further the latent period between earliest exposure and effect manifestation should be as expected on the known biology of the disease in question. Protopathic bias; concluding causation when disease process precedes the risk factor of occurrence, can result if appropriate care is not taken.
5. **Dose response relationship**: The more intense the exposure the greater the risk of disease however a dose response relationship is not necessary to infer causation. One of the strongest relationships of the evidence linking lung cancer and tobacco smoking has always been the consistent finding of a dose response relationship between cigarettes smoked and the risk of lung cancer. Such a clear relationship is however not always the case and in the example of benzene. None of the published studies show a dose response relationship to benzene and development of leukemia.
6. **Biologic plausibility**: The proposed cause and effect relationship should always be supplemented by some scientific theory. At instances, epidemiological studies will produce observations which are not plausible at the study time but subsequently will be proven correct. In the benzene example, does the proposition conflict with knowledge of carcinogenesis generally ? No. Chemical exposure, both in animals and humans has been shown to be associated with the development of malignancy. Benzene is capable of inducing cancer in animal experiments.
7. **Experiment:** Experimental design or a randomized control trial is a common technique used to establish the efficacy of a medicinal product. However in-case of above example of benzene, it is not possible to conduct such experiment for ethical reason. For human studies evidences are collected from observations alone. In case of medicines in patients experimental methods are used. Challenge-dechallenge-rechallenge (CDR) is medical testing protocol for statistically establishing validity and efficacy of a medicinal product.
8. **Analogy:** This would include a similarity to some other known cause effect association. If a similar situation exists for other chemicals or related chemicals, the case becomes stronger. In addition to benzene, other aromatic compounds like toluene and xylene are also carcinogenic and provide more support for the benzene proposition.
9. **Coherence of evidence:** This is summing up of all that has appeared above and its integration into a view as to whether a true association exists. If it is so, what is the magnitude gathered from epidemiological and other evidences. Common sense is the most important part of the judgement.

Thus it is important that every researchers need to follow systematic methods in their research. Few commonly used study designs are mentioned below.

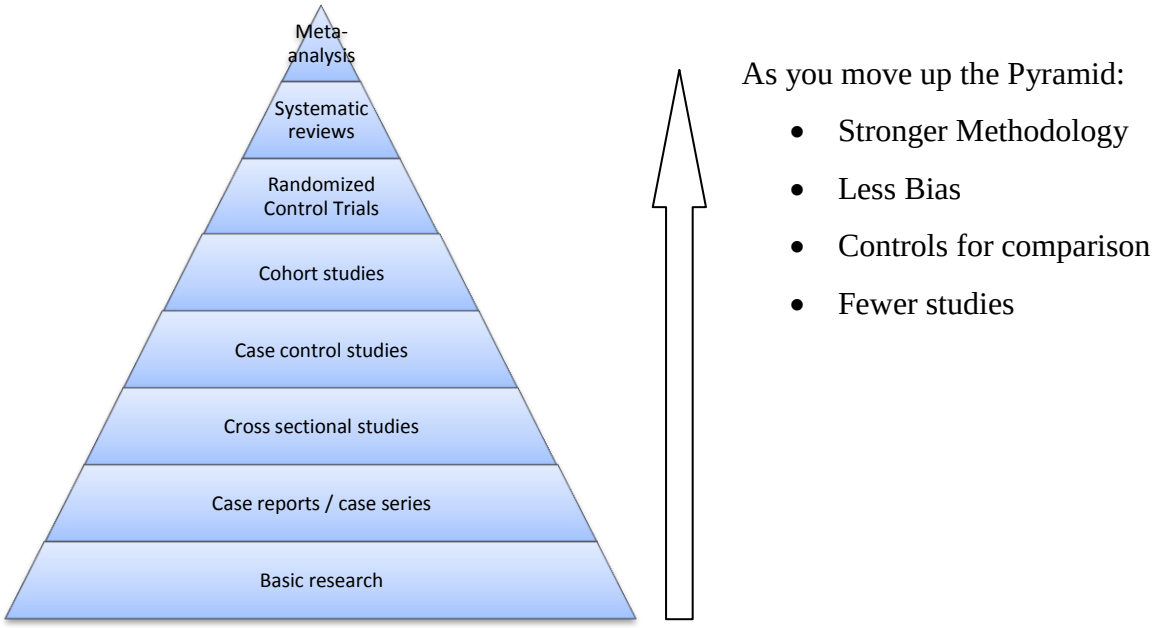
Study Design:

Research design is the conceptual structure within which research is conducted. It is the blueprint for the collection, measurement and analysis of data. Grading of recommendations, assessment, development and evaluation (in short GRADE) working group have developed a common and transparent approach to grading the quality of evidence and strength of recommendations. According to this the following methodologies are ranked high. Type of research question will often suggest the best study design.

For example:

Question	**Study Design**
Clinical examination	Prospective, blind comparison to gold standard
Diagnostic testing	Prospective, blind comparison to gold standard
Prognosis	Cohort study > Case control > Case series
Therapy	RCT / Experimental
Etiology / Harm	RCT > Cohort study > Case control > Case series
Prevention	RAT > Cohort study > Case control > Case series
Cost	Economic analysis

Hierarchy of Research Methods



As you move up the Pyramid:

- Stronger Methodology
- Less Bias
- Controls for comparison
- Fewer studies

Broadly, there are two main types of study design:

I. **Observational study**
II. **Experimental or Intervention study**

I. Observational Studies:

Observational studies are those in which the investigator simply observes and measures the parameters of interest and does not intervene in any of the process.

These can be further categorized into ***descriptive and analytical studies***

1. Descriptive Studies:

A descriptive study is "concerned with and designed only to describe the existing distribution of the variables in the population under study, without regard to causal or other hypotheses." Descriptive studies describe the health conditions and health-related characteristics of populations, typically in terms of person, place, and time. The data may be collected by actual interviews of the subjects, by obtaining measurements of physical characteristics, or by extracting information from existing sources, such as disease registries, hospital or employment records. This provides essential contextual information to develop hypotheses, design studies, and interpret results.

The different types of descriptive studies are ***case reports, case series, ecological and cross sectional studies***.

a. **Case reports** and case series are prevalent in clinical practice and focus the infrequent and rare diseases and/or conditions.
b. **Ecological study** refers to measurement of variable(s) at group or aggregate level (a family, clan or school) or an ecological unit (a village, town or country), rather than individual level and individual level data are missing and not considered for analysis.
c. **Cross sectional study** measures or documents the condition of interest in the population of interest within a specific time period at one point of time. The usual output of this study methodology is prevalence. At one point in time the subjects are assessed to determine whether they were exposed to the relevant agent/ exposure and whether they have the outcome of interest/ disease.

2. Analytical Studies:

Analytic epidemiology involves the systematic evaluation of suspected relationships between the exposure and outcome between at least two groups. Analytic studies are generally adopted to test one or more specific hypotheses, typically whether an exposure is a risk factor for a disease or an intervention is effective in preventing or curing disease (or any other occurrence or condition of interest). The different types of analytical studies are ***case-control, cross-sectional and cohort studies***.

a. **Case Control studies:** In case-control study, the subjects are selected based on their outcome of interest (or disease) status: i.e., cases with outcome of interest and controls without outcome of interest. Data on the exposure of interest is collected retrospectively (in to past) and two groups are compared with respect to prior exposure. Thus, the subjects are selected and sampled based on their disease or outcome status. Although, case-control studies are

usual in infectious disease epidemiology, usage is also common in chronic disease epidemiology. The key point of case-control study is basis of selection of subjects and directionality of enquiry. Case-control studies are important and may be only feasible methodology for diseases with rare outcomes or diseases. The relationship between exposure and outcome in two groups (cases and controls) are expressed as odds ratios. The important part of subject selection is selection of the controls matching with the cases.

b. **Cohort studies:** cohort study refers to prospective follow up of a group of subjects who share common characteristic of interest to document the exposure and outcome of interest over time. Cohort refers to a group of people who share a common characteristic or exposure within a defined period, like birth cohort, occupational cohort, etc. Cohort study is usually prospective, but it may be retrospective, but always the direction of enquiry and data collection is forward. This is the best method for determining the incidence and natural history of a condition. The common strategy of cohort studies is to start with a reference population (or a representative sample thereof), some of whom have certain characteristics or attributes relevant to the study (exposed group), with others who do not have those characteristics (unexposed group). Both groups should, at the outset of the study, be free from the condition or conditions under consideration. Both groups are then observed over a specified period to find out the risk each group has of developing the condition(s) of interest. The data is collected multiple times during the period of observation for all subjects. This is referred as the gold standard observational study.

II. Experimental Studies:

Experimental study is the best design to prove causation and association. It can be viewed as the final or definitive step in the research process, a mechanism for confirming or rejecting the validity of ideas, assumptions, postulates and hypotheses. In these studies, the investigator controls selection of the subjects, the exposure/ intervention, outcome measurements, and sets the conditions under which the experiment is conducted. The selection of subjects is done in such a way that the comparison of outcome measure between the exposed and unexposed groups is as free of bias as possible. The experimental study designs include controlled trials (clinical trials or community intervention trials) and quasi-experimental designs. Clinical trials usually refer to Randomized Clinical Trials (RCTs), although it may be non-randomized also, if the allocation is no random. In research practice, non-randomized trials are not preferred.

a. Randomized clinical trials (RCTs)

Clinical trial in which the subjects after selection are allocated into the experiment or comparison (or placebo) group randomly following a systematic procedure. The importance of random allocation is to balance the both groups so that the only difference is the intervention. Randomization is done to avoid confounding. Random allocation can be done following any of these methods; simple randomization, random number table, computer generated allocation or IVRS (interactive voice response system). Blinding ensures balancing the intervention and comparison groups for the unknown factors and is considered as an essential part of RCTs. Blinding may be single blind (hiding the identity of the intervention from the subject) or double blind (hiding from the subject and investigator) and triple blind (hiding from subject, investigator and also the person who enters and analyses the data). Clinical trials may be categorized according to the purposes; prophylactic trials

(immunization, contraception); therapeutic trials (drug treatment, surgical procedure); safety trials (side-effects of oral contraceptives and injectables); and risk-factor trials (proving the etiology of a disease by inducing it with the putative agent in animals, or withdrawing the agent through cessation). Traditionally clinical trials of new therapies or devices pass through four phases; (1) Phase I trials assessing safety in volunteers; (2) Phase II trials assessing the effectiveness in patients to determine the appropriate dosage, and to investigate its safety; (3) Phase III (referred as the classical clinical trial) assessing the effectiveness and safety of in a larger number of subjects/ patients (this phase findings lead to licensure of the drug or device for use); (4) Phase IV trial (post-marketing trial) re-assessing the effectiveness, safety, acceptability and continued use of the drugs or devices under real clinical conditions. Design and conduct of clinical trials is a specialized activity and resource intensive one. The regulatory and ethical issues are also considered and followed for these RCT studies.

b. Community intervention trials (CITs)

Usually carried out in community, hospitals or clinics, and are usually directed at a subject group with specific health conditions. However, randomized experiments are also sometimes done in the community (vaccine trials). Randomization process is also adopted for community trials, cluster randomization, where communities are randomly assigned to either intervention or comparison group. The major difference from the RCT is that the randomization is done on communities rather than individuals. It should be ensured that the communities selected for the study are similar as much as is possible. Binding may or may not be possible, depending on the intervention. Risk of contamination may be there, when individuals from one group receives the other intervention.

c. Quasi-Experimental Study:

This design falls between observational and true experimental study. This study design does not meet all requirements necessary for controlling influences of extraneous variables and often, random assignment of participants is not possible. In such cases the intervention is often not completely planned by the investigator. This study design is usually used for health system research in which intervention area experiences new or modified program, while the comparison area continues prior program. A "natural experiment" is also similar, but refers to naturally occurring events (e.g., a study of mental health following an earthquake).

III. Other Study Designs

a. Systematic Review

It is a process of comprehensive evaluation and synthesis of all available evidences (published and may be unpublished) for a specific focused question (drawn from research and other source), including appraisal of the methodology, results, key findings, identifying the possible reasons for different results across studies and limitations of current knowledge. A review earns the adjective systematic if it is based on a clearly formulated question, identifies relevant studies, appraises their quality and summarizes the evidence by use of explicit methodology. Cochrane Reviews are examples of Systematic reviews.

b. Meta-analysis

This is a subset of systematic review, which uses statistical techniques for combining the findings from independent studies (a process sometimes referred to as pooling). Meta-analysis is most often used to assess the clinical effectiveness of healthcare interventions by combining data from two or more randomized control trials. Meta-analysis of trials provides a precise estimate of treatment effect, giving due weight to the size of the different studies included and allow taking evidence based decision.

Sample size

Research question and to test a specific hypothesis, the study must include appropriate power, i.e., subjects of minimum number, known as sample size. Since the study cannot include all the target population, a representative subset of population is carefully and systematically chosen (defined as sample). If the sample is too small, the results will not be reliable. On the other hand, if the sample is too large, it prolongs the study and makes it more expensive, with no added scientific value. There are specific equations for estimating sample size for different study designs. Different sample size estimation software (Epi-info, nMaster, PS-power, etc) are also available for this purpose.

Sampling procedures

Sampling indicates the technique of selection of the sample subjects from the population. The sampling technique may be probability sampling or non-probability sampling.

1. **Probability Sampling** is one in which every unit in the population has a chance (greater than zero) of being selected in the sample and minimizes selection bias.

2. **Non-probability Sampling** is any sampling method where some elements of the population have *no* chance of selection or where the probability of selection can't be accurately determined. It involves the selection of sample based on assumptions regarding the population of interest. This technique usually suffers from selection bias and challenge of generalisability.

The common probability sampling techniques are simple random sampling, systematic sampling, stratified sampling, probability-proportional-to-size sampling and cluster sampling. The common non-probability sampling techniques are convenience sampling, quota sampling, snowball sampling and purposive sampling.

Collection of Data

Collection of data is an essential part of the research process. Data can be primary or secondary. Data collected by the researcher, say by a survey, is primary. The data already collected earlier and/or available in some published form is secondary.

For primary data collection, an important question is how data is collected and/or measurements are done to ensure reliability and validity. Reliability means that the observer repeating the test, or someone else using the same method should be able to obtain the same findings. Validity means that the measurement should actually represent what it is intended to measure. To ensure reliability or reproducibility of the results the following should be considered.

- Measurements made should not vary by observer or between observers (intra- and inter-observer consistency).
- Instrument or laboratory variability should be taken into consideration.
- Subject variability should be considered if measurements vary according to the time they are made, for example, fasting or after meal, time of the day, or day of the menstrual cycle.

Obtaining the same result by the same and different raters ensures reliability and reproducibility, but does not mean validity. The test, itself, may not be accurate in measuring what it is intended to measure.

Questionnaire designing

Study questionnaire is a document designed for the purpose of seeking specific information of interest from the respondents. The questionnaire may be self-administered or administered by interviewers. There are two major question formats: the closed-response (with a list of pre-determined response options) and open-ended (without pre-determined response options) types, but a questionnaire may include both question formats. The questionnaire should always be pre-tested in a pilot study before the main survey. Interviewers should be trained to make sure that the questionnaire is administered in a uniform way.

Analysis of data

Analysis involves steps like categorisation, coding, tabulation, etc. The principle for classification or categorisation of data has to be based on the problem under study or the hypothesis formulated. The category must be exhaustive and sufficient for classifying all responses. They must be distinct, separate and mutually exclusive. Coding involves grouping of responses falling under a particular category. Tabulation is a means of organising the responses to facilitate comparisons bringing up the inherent relations between two or more variables. It is an orderly arrangement of data in columns and rows. Analysis and inference is usually aided by the application of different statistical and econometric techniques.

Report writing

Originality and clarity are the two vital components of research report. It is the ultimate test of one's analytical ability and communication skills. It is an exercise involving the organisation of ideas. The research report need to be presented in such a manner that the readers can grasp the context, methodology and findings easily. The report comprise of two parts: the preliminary pages and the main text. In the preliminary pages, the report should indicate the title of the research study, name of the researcher (and his team members) and the name of the institution and/or the month/year of preparation of the report. This should be followed by a 'preface' in which the main context of preparing the report along with key findings must be presented. Towards the end of the 'preface', the important sources/persons can be acknowledged suitably.

The main text begins with an introductory chapter followed by the major aspects of the study organised into different chapters. The introductory chapter should contain a clear statement of the

objectives of the study, rationale behind the study, a brief summary of the literature review, hypotheses tested (if any) and the definitions of the major concepts employed in the study. The methodology adopted in conducting the study must also be fully explained along with an explicit mention of the limitations of the study. The subsequent parts of the main text, should present the major aspects of the study arranged in a logical sequence split into appropriate sections and subsections. The inter-connection between different sections should be properly maintained so that the report reads smoothly.

The report may also include an 'executive summary' outlining the context and methodology, and major findings of the study. The 'executive summary' is placed right at the beginning (i.e. before the introductory chapter) so as to provide a concise picture of the entire report.

Qualitative Research

Most of the above section dealt with planning quantitative research. Qualitative research needs a different approach. Qualitative research focuses on exploring the issues that are not explained by quantitative method. While designing qualitative research, it is important to develop a conceptual framework based on the available literature and probable relationships and/or processes. It helps to outline the research questions, and provides a context for understanding the research.

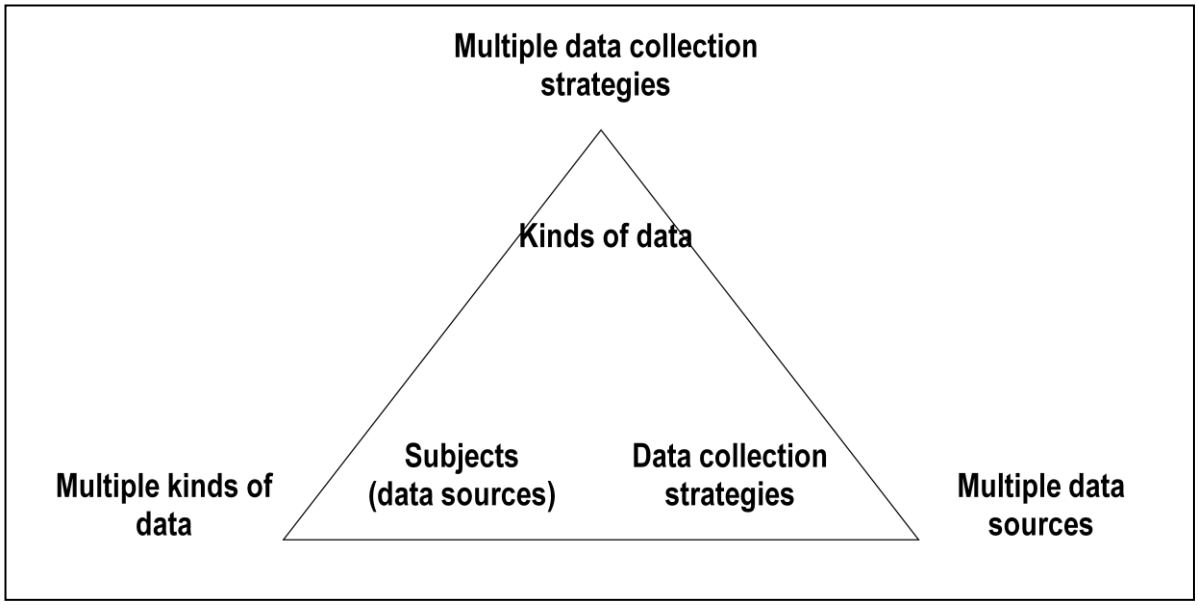
Criterion for qualitative research methods

Often in qualitative research three key criterions are used.

1. **Holistic description**
   When conducting qualitative research, the investigator seeks to gain a total or complete picture. According to Stainback and Stainback (1988) a holistic description of events, procedures, and philosophies occurring in natural settings is often needed to make accurate situational decisions. This differs from quantitative research in which selected, pre-defined variables are studied.

2. **Corroboration**
   The purpose of corroboration is not to confirm whether people's perceptions are accurate or true reflections of a situation but rather to ensure that the research findings accurately reflect people's perceptions, whatever they may be. The purpose of corroboration is to help researchers increase their understanding of the probability that their findings will be seen as credible or worthy of consideration by others (Stainback & Stainback, 1988).

3. **Triangulation**
   This is a method to enhance the validity & reliability of qualitative research. It enhances accuracy of interpretation and confirms that the data collected is not due to chance or circum-stances. Often triangulation is represented as below.



Three main methods are commonly used in qualitative research: observation, in-depth interviews (IDIs) and focus group discussion (FGD). The investigator has to select which method would be more appropriate to answer the research question, or may use more than one method. The researcher in these different designs plays the role of observer (for observation), interviewer (IDIs) or group moderator (FGDs).

1. Observation

Observation can be made from an outsider or insider perspective, or somewhere in between, depending on the objective of the study. As an outsider, researcher collects data being an observer from a distance (observing the quality of health care delivery in a clinic, health centre or a pharmacy), but as an insider, researcher interacts with participants for data collection. "Time and motion study" is a special type of observation study, to document how the subject functions or performs duty and uses time. The researcher observes what a health worker is doing over a defined.

2. In-depth interviews

In-depth interviews are intensive one-on-one interviews in a more social manner without any structured questionnaire. Open-ended questioning is a basic tool in qualitative research. Although there may be a topic guide for the IDIs, but the questions flow from the answer of the respondent, as a follow-up to the answer, or to probe further into the answer. In case of multisite studies, a pre-determined set of open-ended questions are the most standardized approach for IDIs.

3. Focus groups

For obtaining additional information from the interaction of a group of respondents, focus group discussions are used. The FGDs complement IDIs and add to the information and insights gained from the IDIs. FGD is practically exchange of information, ideas and views among the participants themselves and the researcher plays the role of a moderator, and not an interviewer. For effective FGD, the group should be relatively homogeneous, (e.g. age, sex and sociocultural background) depending on the research question. Usually a group of 8-12 participants are adequate for a good and

manageable discussion. One or more rapporteurs or note-takers record what people say, document the sociogram (process and flow of interaction and exchanges in a pictorial manner), but also the body languages of participants.

Analysis of Qualitative Data:

Analysis if qualitative data is different from quantitative ones. Different processes are adopted for analysis including free listing of responses, domain identification, coding followed by cross tabulation according to the codes and summarizing. Although several softwares are available for qualitative data analysis, many researchers prefer to do it manually.

Writing a Research Protocol

Once we have chosen an appropriate study design, determined our method of treatment allocation, and calculated the sample size necessary to identify a significant result, it is time to perform the study "on paper". Creation of a research protocol allows us to carefully consider and define each step of the study, anticipating potential problems and sources of error before the actual data collection begins. It is also a necessary first step in applying for research funding and ethical approval. A typical research protocol considers the elements listed below, each of which will be addressed in detail.

- Purpose
- Patient population
- Study hypotheses
- Methods (including treatment)
- Inclusion and exclusion criteria
- Outcome variables
- Data to be collected
- Statistical methods to be applied
- Anticipated economic cost

Purpose: Any research protocol should begin by considering the question "Is this study necessary and, if so, why?" This is especially true in this day and age where research funding is becoming more scarce, competition for funding more fierce, and IRB approval more difficult to obtain. The rationale for performing the study should be clearly defined and previous studies in the medical literature which support the need for this study identified. The scientific background for the study should be outlined such that a grant reviewer or IRB committee, upon reading the protocol, will have sufficient knowledge to understand the study and its hypotheses.

Study Hypotheses: The primary research hypothesis and all secondary hypotheses should be clearly stated. Some studies will only have a single hypothesis while others will have several. It is generally a good idea to limit the number of hypotheses in a study to a primary hypothesis and no more than 4 or 5 secondary hypotheses. Multiple hypotheses can result in an unwieldy study that is complicated and therefore difficult to perform accurately. Simplicity is best.

Methods: This is the most important part of the research protocol. It is here that the details of the study are defined and potential problems identified. All necessary equipment and supplies should be listed as well as the manner in which data will be measured and collected. The treatment being

administered, if any, should also be clearly described including the method by which treatment will be allocated.

Patient Population: The patient population of interest should be described in detail to provide a clear image of to which population the study conclusions will be applicable. This is also an essential step in ensuring that the method of treatment allocation will result in a representative patient sample being chosen.

Inclusion and Exclusion Criteria: The criteria which will be used to identify patients eligible for the study ("inclusion criteria") should be identified as well as those criteria which will eliminate a patient from being in the study ("exclusion criteria"). It should be kept in mind that the more specific the inclusion criteria, the smaller the population to which the results will be applicable. For example, if we include "all adult males" in our study, the conclusions that we make will be applicable to a large patient population. If, however, we limit the study to "all males between 18 and 25 years who have hypertension," the population to which our conclusions will be relevant is a much smaller one. We must therefore take care to ensure that our inclusion criteria include the population to which we wish to apply our results.

Our choice of exclusion criteria is just as important, if not more so. Exclusion criteria represent those criteria by which we will reject patients from entering the study who otherwise meet the entrance criteria. It is essential that the exclusion criteria be defined prior to and followed carefully during the study in order to prevent the introduction of selection and assembly bias.

Outcome Variables: Those variables which will be used to define response to the treatment (or lack thereof) should be identified. These variables might include survival, presence of organ failure, successful extubation, decreased length of stay, or any of a multitude of other outcome measures.

Data to be collected: All of the independent variables to be evaluated during the study should be listed. It is important to consider, before the study begins, which types of data will be necessary to answer the study hypotheses. It is frequently difficult, if not impossible, to obtain or recreate data after data collection is completed. Care should therefore be taken to anticipate and collect all of the data that may be required during data analysis. It is always better to have too much data than too little. By the same token, data collection should be limited to those variables which are pertinent to the study hypotheses. The collection of extraneous data is time consuming and costly, and tends to decrease the validity of the data collection overall.

Statistical Analysis: Considering the types of data variables which are being studied, it is generally a good idea to list the statistical methods that will be used during data analysis. By considering the statistical analysis before the study, the need for inclusion of further data variables may become apparent.

Budget: All costs that will be incurred by performance of the study should be listed including labor and equipment costs, pharmacy costs, laboratory fees, radiologic procedure charges, etc. Care should also be taken to account for unforeseen costs that may arise once the study begins.

There are several benefits of taking the time to prepare a research protocol prior to beginning a study. First, as previously mentioned, it allows us to perform the study "on paper" in such a manner that we can anticipate and correct for problems, potential errors, and sources of statistical bias before the actual study begins. Second, it is a necessary requirement for applying for research funding or IRB

approval. Third, having drafted the research protocol, the preparation of abstracts and manuscripts for publication is simplified. The background literature research has already been performed and the Introduction and Methods sections have essentially been written. Thus, as we will see in the next chapter, the abstract and manuscript for a study have largely been composed before the study is ever performed.

Ethics in Research:

A number of developments have brought the subject of ethics in medical research to the front line of concerns of the health profession and the society at large. These include a major expansion in health research, the significant public investment in research, the increasing need for experimentation on human subjects, publicized cases of ethical violation, internationalization of research, and the expanding role of private industry.

General Ethical Principles:

Where research involves experimentation on human subjects, every effort should be made to maximize the benefits to the subjects (beneficence), and the subjects should suffer no harm (non-maleficence). The principle of respect implies that participation in the research should be completely voluntary and based on informed consent. Where research involves collection of data on individuals, privacy should be protected by ensuring confidentiality. Respect to the community means respecting its values and having its approval for the research. The principle of justice (distributive justice) implies that participation in the research should correlate with expected benefits. No population group should carry an undue burden of research for the benefit of another group.

Apart from the basic principles of beneficence (non-maleficence), respect and distributive justice, other principles also apply. Where research involves experimentation on animals, mercy is an ethical imperative. For research in general, medical or non- medical, honesty is an indispensable value. International ethical guidelines for biomedical research involving human subjects have been issued by the Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization. The latest edition was issued in 2000.

Responsibility for ensuring that ethical standards are observed in research rests collectively with the investigators, research institutions, national drug regulatory agencies, editors of medical journals, and funding agencies and organizations. Ethical approval by one does not relieve the others of responsibility.

Countries and institutions should establish ethical review systems to ensure the protection of potential research participants and contribute to the highest attainable quality in the science and ethics of health research. Ethics committees should be established, as appropriate, at the national, regional and institutional levels.

The World Health Organization has issued operational guidelines for ethics committees that review biomedical research, outlining their role, how they can be constituted, procedure for submitting an application, elements for review, decision- making, follow-up, and documentation and archiving (WHO, 2000). The elements of ethical review include scientific design and conduct of the study, recruitment, care and protection of research participants, protection of participant confidentiality, informed consent process and community considerations. Some aspects of the work of ethics committees need to be highlighted.

Ethical considerations apply throughout the research process, and shall be taken as integral components of the research process. It is unethical to expose subjects to research that is not scientifically sound, is not performed by qualified investigators in qualified facilities, and is not likely to provide valid scientific answers. In fact, scientific assessment of the planned research is an important part of the ethical review process.

References & suggested readings:

1. Hauser. M et.al., Conducting bio-social surveys: collecting, storing, accessing and protecting bio-specimens and bio-data. National Research Concil. (2010) [Downloaded from : http://www.nap.edu/catalog/12942.html]

2. University of Ottawa. Society, the Individual and Medicine. [Website article, accessed on 21-01-2015, from: http://www.med.uottawa.ca/sim/data/Study_Designs_e.htm

3. Ann Aschengrau, Seage GR. Essential of Epidemiology in Public Health. 3rd edition

4. Sage publications. Introduction to hypothesis testing. [Website article, accessed on 31-01-2015, from: http://www.sagepub.com/upm-data/40007_Chapter8.pdf]

5. Type of samples [Website article, accessed on 21-01-2015, from: http://psychology.ucdavis.edu/sommerb/sommerdemo/sampling/types.htm

6. Ethics in health research. [Website article, accessed on 21-01-2015, from: http://whqlibdoc.who.int/emro/2004/9290213639_chap2.pdf

7. Planning the research. [Website article, accessed on 21-01-2015, from: http://whqlibdoc.who.int/emro/2004/9290213639_chap4.pdf