"Basic Course in Bio-medical Research"

Online course for post-graduate medical students and faculty

Course approved by Board of Governors - in supersession of Medical Council of India
In partnership with ICMR-National Institute of Epidemiology and SWAYAM NPTEL

Basic Course in Biomedical Research – Review

Dr. Amit M Shah
Associate Professor
Department of Pharmacology
GMERS Medical College
Gandhinagar
M: 9974820709

This book contains the compilation of PowerPoint presentation of all 23 lectures of 'Basic course in Biomedical Research' along with Assignments (with correct answers) of Cycle 1 (Sep-Dec 2019) and Cycle 2 (Mar-Jun 2020). MCQs no. 1 to 10 belong to Cycle 1 and MCQs no. 11 to 20 belong to Cycle 2. I have just compiled all this information for ease of use. I don't owe any copyright on the material present in this e-book. It is strictly for study purpose not for the commercial purpose. It can be helpful for all the faculties/students for completion of 'Basic course in Biomedical Research'.

Contents

Dimension of health research

- Theoretical research and applied research
- Preventive and therapeutic research
- Bench based research and bedside research
- Exploratory research and confirmatory research
- Implementation research and translational research

Fundamental principles to be followed

- Planning stage is very critical – spend enough time and involve the right people in planning
- Team work is critical
- Three levels of review are essential
  - Scientific review: novelty, rationality, justification
  - Ethics review: human subjects protection
  - Regulatory review: foreign funding, sample shipment, intellectual property, exchange of visitors

Process of health research

- Ensure that data is collected systematically
- Draw meaningful conclusions
- Make appropriate decisions
- Take appropriate actions for prevention and control of diseases, conditions: Evidence based actions
- These should help in reduction of suffering and ultimately improve health and well-being of the community

Breadth and depth of inquiry in health research

- Human host: healthy, susceptible, with disease, dead
- Surrounding environment and society: climatic factors, housing, vectors, animals, socio-cultural practices, family structure
- Health care infrastructure and delivery

Broad scope of health research

- Getting additional or new information
  - Are more of diphtheria and pertussis reported among adults in recent times?
  - What are the differences in full genome structure of HBV and HEV?
- Verifying and confirming available information
  - Are etiologies of pediatric pneumonia different in the children aged 5 or less in developed and resource limited countries?
  - Have the incidence and complications of diabetes changed with increasing consumption of pre-cooked and packaged food?
- Explaining cause and effect relationship
  - Does presence of a particular co-receptor [cause] on CD4 cells protect against HIV infection [effect]?
  - Are breast cancers [effect] more common in breast implant [cause] recipients?

Broad scope of health research

- Testing new drugs, vaccines, tools or interventions for prevention, treatment and control of a disease
  - Can INH prophylaxis delay onset of tuberculosis in HIV infected persons?
  - Will introduction of smokeless stoves result in reduction of respiratory morbidity and mortality in rural areas?
- Evaluating ongoing programs and assessing feasibility of new programs
  - Is injectable iron sucrose a better alternative to deal with pregnancy related anemia than oral iron?
  - Will the Integrated Disease Surveillance Program be able predicting the epidemics of influenza and bird flu in India?

Making the right choice of study design

- Qualitative studies or Quantitative studies
- Observational studies or Experimental studies
- Retrospective studies or Prospective studies

Some critical considerations in planning phase

- There should be adequate justification to conduct the research study
- The research question should have clarity and focus
- Case definitions of study variables and outcomes should be standard and unambiguous
- Sample and sample size:
  - Should be representative of the population [External validity or generalizability]
  - Should be adequate [power to draw meaningful inferences]

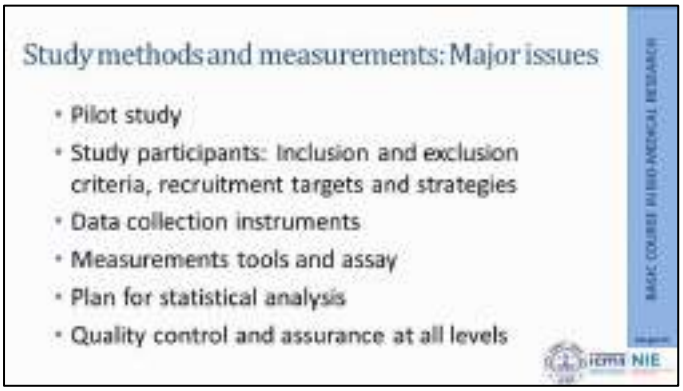
Research can never be free of errors, but errors can be predicted and minimized

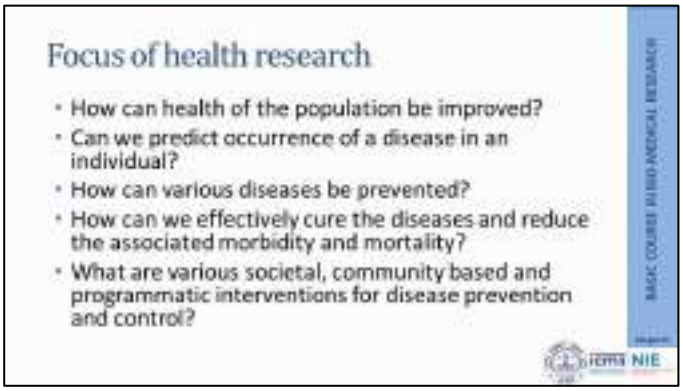
- Random error representing wrong result due to chance: unknown sources of variation that can distort findings in either direction
  - Can be minimized by increasing sample size and increasing precision
- Systematic error signifying wrong result due to bias - mostly due to variation that would distort the results in one direction
  - Can be minimized by improving study design

Challenges in designing and implementation of research studies

In a scenario when we desire to study the relationship between a variable and an outcome

- Confounders: Affect both study variable as well as outcome
  - Effect can be minimized by proper study design and through stratified analysis
- Effect modifiers: Can alter [generally negatively] the relationship between the study variable and outcome by independently affecting outcome
  - Good to be aware of them through adequate literature review and not to include them in the study





1. Which of the following areas can be included in health research?
   i. Improving the health of the population.
   ii. Predicting progression of a disease in a patient
   iii. Prevention of various diseases
   iv. To explore various societal, community based and programmatic interventions for disease prevention and control
   a) i and ii
   b) i, ii and iv
   c) All of the above
   d) None of the above

2. What is appropriate for sample and sample size?
   a) Should be representative of the population [External validity or generalizability]
   b) Should be adequate [power to draw meaningful inferences]
   c) Both 'a' and 'b'
   d) Neither 'a' nor 'b'

3. Which of the following statements are correct regarding defining inclusion and exclusion criteria in a study protocol?
   a) They should be vague because this will allow greater and easy enrollment
   b) They should be very specific
   c) They should be very large in number
   d) It is not important to define exclusion criteria in a clinical trial

4. Which of the following are examined as part of regulatory review?
   a) Information regarding transfer of funds and utilization of funds
   b) Shipment of samples and transfer of data outside the country
   c) Sharing and protection of intellectual property
   d) All of the above

5. Which of the following statements is not true in case of pilot study?
   a) They are conducted for developing and testing adequacy of research instruments
   b) They establish whether the sampling frame and technique are effective
   c) Ethics committee approves the main study only after successful completion of the pilot study
   d) They are small scale studies

6. Before initiating a study involving primary data collection, the Principal Investigator must ensure that various approvals are obtained. Which of the following approvals is absolutely mandatory?
   a) Scientific committee approval
   b) Ethics committee approval
   c) Technical committee approval
   d) Regulatory authority approval

7. Which is the best source of information on 'effect modifiers' while exploring cause and effect relationship in a research study?
   a) Deductive thinking
   b) Thorough review of literature
   c) Intelligent guessing
   d) Discussing with experienced researchers

8. The policy makers want to know whether introduction of pentavalent vaccine in the national program is resulting in reduction in the number of Hemophilus

influenza cases. Which of the following studies will they have to conduct to find an answer?
a) Case-control study
b) Field trial
c) Ecological study
d) Case series

9. What is true about Confounders?
a) They affect both study variable as well as outcome
b) Their effect can be minimized by proper study design and through stratified analysis
c) Both 'a' and 'b'
d) None of the above

10. Which of the following is not a type of study design?
a) Qualitative study
b) Observational study
c) Retrospective study
d) Pilot study

11. Any systematic error in the design, conduct or analysis of a study that results in an erroneous estimate of an exposure's effect on the risk of disease is called:
a) Confounding
b) Bias
c) Interaction
d) Stratification

12. Which of the following is not part of ethics review of a project?
a) Informed consent document and procedure
b) Competence of researcher and institute conducting research
c) Sharing and protection of intellectual property
d) Care and support during and after completion of research

13. A study was conducted to assess the extrapyramidal side effects of a new antipsychotic drug in patients with schizophrenia. Many of these patients were smokers and some of them were on anticholinergic drugs. What was the role of the anticholinergic drugs in this study?
a) Confounder
b) Random Variable
c) Effect Modifier
d) Independent Variable

14. Before initiating the study, the Principal Investigator must ensure that various approvals are obtained. Which of the following approvals is/are absolutely mandatory?
a) Scientific committee approval
b) Ethics committee approval
c) Technical committee approval
d) Regulatory authority approval
e) All

15. The policy makers want to know whether introduction of a new rotavirus vaccine in the national immunization programme is resulting in reduction of morbidity and mortality from rotavirus disease. Which of the following studies will they have to conduct to find an answer?
a) Case-control study
b) Ecological study
c) Field randomized trial
d) Case-series

16. What effect does increasing the sample size have upon the random error?
a) It increases the random error
b) It has no effect on the random error
c) It reduces the random error
d) None of the above

17. Which of the following will best describe the scientific inquiry that seeks to understand the acceptability and functionality of a health program?
a) Basic science research
b) Translational research
c) Clinical research
d) Implementation research
e) None of the above

18. The following statements describe confounding and effect modification. Which of the statement is/are correct?
a) In a study of relationship between coffee drinking and oro-pharyngeal cancer; smoking is a confounder
b) In a study to explore relationship between hepatitis B infection and post-infection hepatic sequelae, habit of alcohol drinking acts as an effect modifier and patients with this habit may be excluded from the study
c) 'a' and 'b' are correct
d) 'a' and 'b' are wrong

19. Issues regarding shipment of samples and transfer of data outside the country are examined by:
a) Regulatory review
b) Ethics review
c) Scientific review
d) None

20. Which of the following disciplines contribute to health research?
a) Bio-medical research
b) Biostatistics
c) Social science research
d) All of the above

Key areas

- Spell out research question
- State research hypothesis
- Formulate objectives

The life cycle of research

What is research question?

- 'Uncertainty' about something in the population that the investigator wants to resolve by making measurements in the study population
- Uncertainty = 'data needs'
- Clear question facilitates to
  - Choose the most optimal design
  - Identify who should be included, what the outcomes should be, and when the outcomes need to be measured

Refining 'ideas' into research questions

- Begins with general uncertainty about a health issue
- Narrows down to a concrete, researchable issue

Translating uncertainty to research question

- Frames problem in specific terms (clinical/public health/...)
- Focuses on one issue
- Is written in everyday language
- Can use more than one operational verb, if needed
- Should link the question to the potential action that would be taken once the question is answered
- *Is stated as a question!*

Research question sets out

- ✓ What the investigator wants to *know*
- ☒ NOT
  - ☒ What the investigator might *do* or
  - ☒ What the results of the study might ultimately *contribute* to that particular field of science

Sources of research questions

1. Mastering the published literature
   - Continue review of work of others in the area of interest
2. Being alert to new ideas and techniques
   - Attending research meetings / conferences
   - Having a skeptical attitude about prevailing beliefs
   - Applying new technologies to old issues
3. Keeping the imagination roaming
   - Careful observation; teaching, tenacity
4. Choosing a guide/mentor

SB Hulley et al. Designing Clinical Research, 3rd ed. Lippincott Williams & Wilkins 2007

Two categories of research questions

1. Descriptive questions

- Involve observations to measure quantity
- No comparison groups / interventions

2. Analytical questions

- Involve comparisons / interventions to test a hypothesis

Steps in conceiving a research question

1. Review of state-of-art information
2. Raise a question
3. Decide worth investigating by peer-review
4. Define measurable exposures & outcomes
5. Sharpen the initial question
6. Refine the question by specifying details

Steps in conceiving a research question

e.g., Should diabetics do exercise daily?

1. Review of state-of-art information

- Exercise reduces blood sugar, body fat
- Exercise improves protection against developing diabetes related complications

Steps in conceiving a research question

2. Raise a question

- *Can exercise help control blood sugar level?*
  - *Rather vague; Need to define*
- *'exercise' & 'blood sugar level'*

Steps in conceiving a research question

3. Decide worth investigating by peer-review

- What is the level of reduction in blood sugar?
  - Fasting or random or post-prandial <i.e., after food>
- What are optimal type, frequency, intensity and duration of exercise?
- What are the risks? What are the other benefits?

Steps in conceiving a research question

4. Define measurable exposures & outcomes

- **Exposure:** Exercise
  - *Pre-determined physical activity comprising of any body movement produced by skeletal muscle, resulting in an increase in energy expenditure*
  - *Atleast one session of 60 minutes every day for one year*
  - Could be specific: *walking, jogging or cycling or aerobic..*
- **Outcome:** Fasting blood sugar level

Steps in conceiving a research question

5. Sharpen the initial question

- *Among diabetics, does physical activity for one hour daily help in reducing fasting blood sugar level?*

Steps in conceiving a research question

6. Refine the question by specifying details

(Study population, operational definitions of variables and study design)

- *What is extent of walking practiced by diabetics (type 2 diabetes) regularly?* [Descriptive question]
- *In order to improve management of type 2 diabetes, we wish to know whether brisk walking by diabetics for atleast one hour daily reduce fasting blood sugar level as compared to those who do not?*

[Analytical question]

Good research question should pass the *'so what?'* test

- **F**easible
- **I**nteresting
- **N**ovel
- **E**thical
- **R**elevant

SB Hulley et al. Designing Clinical Research. 3rd ed. Lippincott Williams & Wilkins 2007

Good research question should pass the *'so what?'* test

- **F**easible
  - Adequate number of participants, technical expertise & resources
- **I**nteresting
- **N**ovel
  - Confirms, refutes or extends previous findings
  - Provides new information
- **E**thical
  - Amenable to a study that ethics committee will approve
- **R**elevant
  - Advance scientific knowledge, improve practice, influence policy

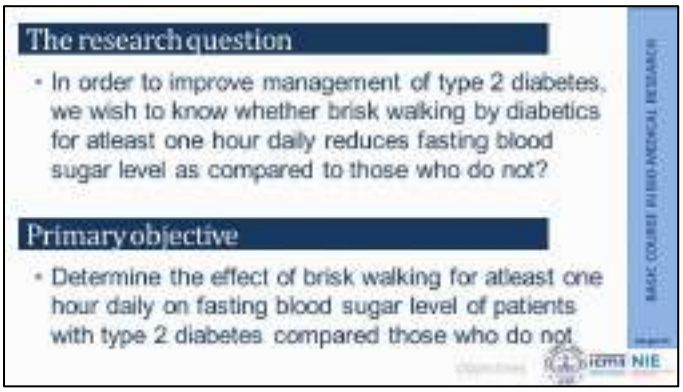
Statement of research hypothesis

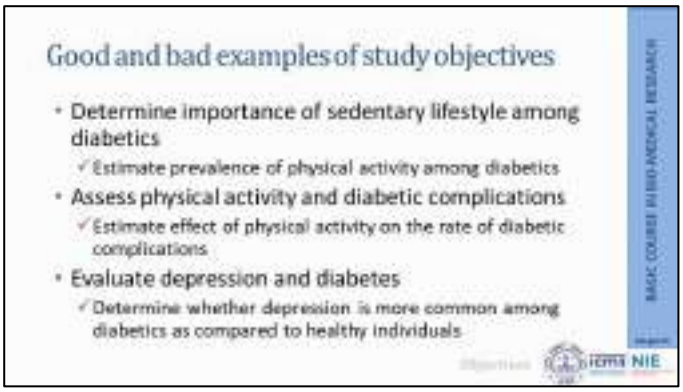
- A specific version of research question
  - Summarizes main elements of study
  - Establishes basis for test(s) of statistical significance
    - Main elements: *Sample, Exposures and Outcomes*
- Stated for analytical questions with comparison groups
  - For research questions with terms: *greater or less than, causes, leads to, compared with, more likely than, associated with, related to, similar to or correlated with*
- Purely descriptive questions DO NOT require hypothesis

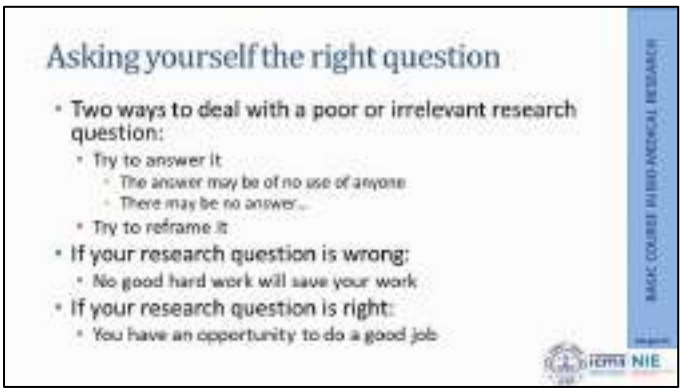
Example of research hypothesis

Among diabetics (type 2 diabetes) from the study area, who do brisk walking for atleast one hour daily results in average reduction of 10 mg% of fasting blood sugar level as compared to those who do not

Characteristics of good hypothesis







1. Which of the following is an element of life cycle of research?
   a) Identify data needs and spell out the research question
   b) Formulate the objective and design the study
   c) Draw conclusion and give recommendation to stakeholders
   d) All of the above

2. Which of the following verbs is preferably used in the statement of objectives of an analytical research study?
   a) Estimate
   b) Determine
   c) Study
   d) Describe

3. Source(s) of research question is/are
   a) Published literature
   b) Being alert to new ideas
   c) Careful observation and teaching
   d) All of the above

4. Which of the following is stated mainly for statistical purpose?
   a) Research question
   b) Objectives
   c) Research hypothesis
   d) All of the above

5. If your objective is to estimate the prevalence of a health problem in a community in 2019, Identify the type of research question this study is addressing
   a) Analytical research question
   b) Descriptive research question
   c) Hypothetical research question
   d) Experimental research question

6. What is the first step in the life cycle of research?
   a) Spell out the research question
   b) Formulate the objective of the study
   c) Identify the data needs
   d) Choose the study design

7. A clear research question facilitates to do the following
   a) Choose the most optimal design
   b) Identify who should be included as study population

c) Specify the outcomes that should be measured
d) All of the above

8. Which of the following statements is incorrect?
a) A good research question should be Feasible, Interesting, Novel, Ethical and Relevant
b) A good research question should be in epidemiological terms
c) A good research question facilitates to choose optimal design
d) A good research question will focus on one issue

9. The verb "estimate" is used in the objective of analytical research studies
a) True
b) False

10. The process of refining the "ideas" into research questions begins with general uncertainty about a health issue and narrow down to a specific, concrete researchable issue
a) True
b) False

11. Which of the following statement is incorrect about a good research question?
a) Research question should advance scientific knowledge, improve practice, influence policy
b) Research question should be approved by the ethics committee
c) Research question should confirms, refutes or extends previous findings
d) Feasibility should not be a criterion while developing research question

12. All the following are characteristic of good research hypothesis EXCEPT
a) Research hypothesis should be simple
b) Research hypothesis should be devoid of any ambiguity about study participants and variables
c) Research hypothesis should be focused on primary objective
d) Research hypothesis should be written once the study is completed

13. Which of the following verb is used in the statement of objective of a descriptive research study?
a) Estimate
b) Determine
c) Examine
d) Compare

14. A research question states about what the results of the study might ultimately contribute to that particular field of science
a) True
b) False

15. Purely descriptive research questions do not require a hypothesis
a) True
b) False

16. Identify the type of research question if the objective of a study is " To determine the effect of tobacco cessation services on tuberculosis treatment outcomes among patients with tuberculosis under National Tuberculosis Elimination Program"?
a) Descriptive research question
b) Hypothetical research question
c) Analytical research question
d) Experimental research question

17. As per the following objectives a hypothesis should be stated for which study?
a) To determine the association of maternal smoking during pregnancy with respiratory infectious disease morbidity and mortality in infants
b) To estimate the lifetime prevalence of mental health morbidities among elderly people in India
c) To describe the pattern of physical activity among school going children aged 6-18 years
d) To describe the temporal and spatial trends of mortality due to cardiovascular diseases, by age and sex in India during 2009-2019

18. All the following are components of 'FINER ' criteria for a research question EXCEPT
a) Feasible
b) Reliable
c) Novel
d) Ethical

19. Which is the last step in the life cycle of research?
a) Spell out the research question
b) Formulate the objective of the study
c) Formulate recommendations
d) Choose the study design

20. The following are the steps in framing a research question.
I. Review of state-of-art information
II. Define measurable exposures & outcomes
III. Raise a question
IV. Decide worth investigating by peer-review
Choose the correct sequence of framing a research question from below.
a) I, II, III, IV
b) II, IV, III, I
c) I, III, IV, II
d) III, I, II, IV

Why perform literature review?

Lit review : Not just a summary

Information seeking
Scan the literature efficiently using manual or computerized methods to identify a set of potentially useful articles and books

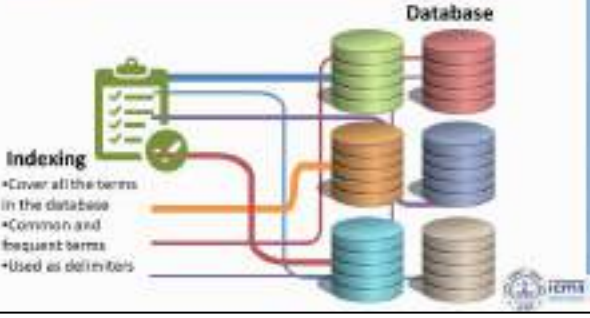
Critical appraisal
the ability to apply principles of analysis to identify those studies which are unbiased and valid

Lit review: Organized search

Information retrieval

identifying, within a large document collection, a subset of documents whose content is most relevant to user's need

Database structure and management



Appropriate place to search

Pubmed (http://www.ncbi.nlm.nih.gov/pubmed)

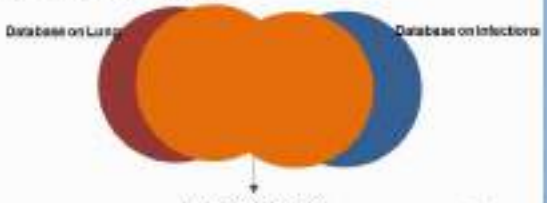
MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.

Searching a database

- Boolean query: AND

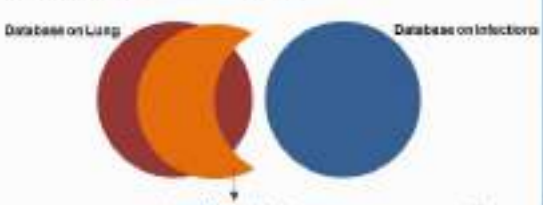
Searching a database

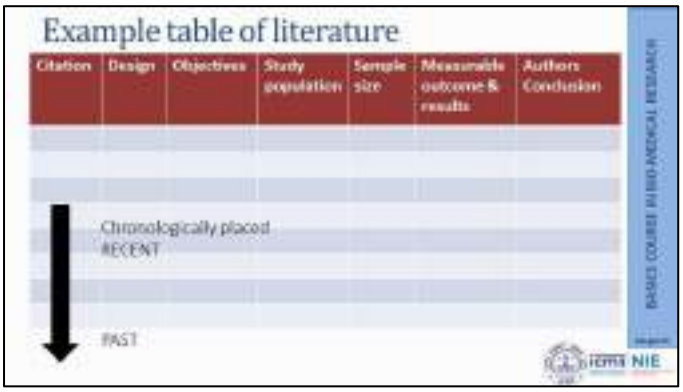
- Boolean query: /OR/



Searching a database

- Boolean query: NOT





1. The ability to apply the principles of analysis to identify those studies which are unbiased and valid is called as
   a) Critical appraisal
   b) Information seeking
   c) Information management
   d) Systematic Review

2. A _____ is a collection of articles, abstracts, scientific proceedings, books, citations etc. that is organized so that it can easily be accessed while doing literature review
   a) Database
   b) Critical appraisal
   c) Hard disk
   d) Index

3. Why should we need to do a literature review?
   a) Save yourself from work
   b) Know the subject matter better
   c) Suggest new research topics, questions and methods
   d) All of the above

4. The process of scanning the literature efficiently using manual or computerized methods to identify a set of potentially useful articles and books is called as
   a) Information seeking
   b) Critical appraisal
   c) Database management
   d) Information retrieval

5. In the Boolean search strategy AND tells that database that you want records that contain all the words you specify
   a) True
   b) False

6. Which of the following is unethical while writing a Literature Review?
   a) The contents from the studies should be presented honestly
   b) The contents from the studies should not be distorted
   c) It is not necessary to address the weakness of the study in a scholarly manner
   d) Sources should be accurately documented

7. Critical appraisal is done in an organized and systematic manner

a) True
b) False

8. The process of identifying, within a large document collection, a subset of documents whose content is most relevant to user's need is called as
a) Information retrieval
b) Information management
c) Systematic Review
d) Narrative Review

9. The query system in the information retrieval process of literature review is
a) User defined
b) Provider defined
c) Conditional
d) Not structured

10. In the National Library of Medicine (NLM), MeSH means
a) Medical Services Heading
b) Medical Subject Heading
c) Medical Subject Helpline
d) Medicine Services Helpline

11. In literature review method of identifying studies which are unbiased and valid is known as critical appraisal.
a) True
b) False

12. Choose the correct sequence of the steps of systematically doing literature search from below
a) Organize the information, identify the lacunae, develop the research question, synthesize the results
b) Identify the lacunae, develop the research question, synthesize the results, organize the information
c) Develop the research question, synthesize the results, organize the information, identify the lacunae
d) Organize information, synthesize the results, identify the lacunae, develop the research question

13. All the following about literature review is correct EXCEPT
a) It identifies lacunae in the existing knowledge about a topic
b) It saves valuable time for a researcher
c) It helps the researcher in arriving the conclusion of a study
d) It suggests the researcher about new research topics

14. While drafting a scientific manuscript, literature review is useful on the following section EXCEPT
a) Introduction
b) Methods
c) Results
d) Discussion

15. Which of the following about PubMed is incorrect?
a) PubMed comprises more than 25 million citations for biomedical literature
b) Citations may include links to full-text article from PubMed Central
c) PubMed is developed and maintained by the National Centre for Biotechnology Information (NCBI), at the U.S. National Library of Medicine (NLM)
d) PubMed is a paid service provider for searching of literature

16. All the following are examples of databases EXCEPT
a) MEDLINE
b) EMBASE
c) CINAHL
d) Google scholar
e) None of the above

17. Literature searches are important to do at the start of a project; and do not need to continue throughout the project.
a) True
b) False

18. Which of the following search query should be used to find the articles on chest pain other than angina?
a) Chest pain AND angina
b) Chest pain OR angina
c) Chest pain NOT angina
d) Chest pain EXCEPT angina

19. A researcher wants to assess effects of polyunsaturated fatty acids (PUFA) on diabetes prognosis. For this, the researcher searches related articles in PubMed and Google scholar. This process is known as
a) Information retrieval
b) Indexing
c) Critical appraisal
d) Data management

20. Which of the following about MeSH is incorrect?
a) MeSH thesaurus is controlled vocabulary produced by the National Library of Medicine
b) It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity
c) It is used for indexing and searching of biomedical and health-related information
d) MeSH is used for EMBASE database

Population at risk

- Portion of a population that is susceptible to a disease
- Can be defined on the basis of demographic or environmental factors

Population at risk: Examples

- Population at risk of developing carcinoma of the cervix:
  - Female population
  - Age > 30 and < 70 years
- Population at risk of hepatitis B
  - Those individuals anti-HBc negative

Prevalence – (P)

- Number of existing cases (old and new) in a defined population at a specified point of time

$$P = \frac{\text{\# people with disease at a specified time}}{\text{Population at risk at the specified time}} \times 10^n$$

- In some studies the total population is used as an approximation if data on population at risk is not available

Point prevalence

- Measures the frequency of disease at a given point in time
- Applies when the data has been collected at one point in time
- P = C / N
  - C = # of observed cases at time 't'
  - N = Population size at time 't'

Example of point prevalence

- 150 children in a school
- Screening for refractory errors at time "t"
- 15 children require glasses
- Prevalence of refractory errors
  - 15 / 150 = 10%

Period prevalence - (PP)

- Measures the frequency of disease over some time
- Applies when the data has been collected over a period of time
- PP = C + I / N
  - C = # of prevalent cases at the beginning of the time period
  - I = # of incident cases that develop during the period
  - N = size of the population for this same time period

Exercise

- Scenario
  - Population of 150 persons
  - Follow up for one year
  - 25 had a disease of interest at the beginning
  - Another 15 new cases developed during the year
- Calculate:
  - Point prevalence at the start of the period
  - Period prevalence for the year

P = C/N = 25 /150 = 0.17 (17 %)

PP = (C+I)/N = (25+15)/150 = 0.27 (27 %)

Factors influencing prevalence

- Number of new cases
- Duration of the illness
  - If the disease is short, the prevalence is reduced
    - The prevalence of sudden infant death = 0
  - If the disease is long, the prevalence is increased
    - Rare lifelong disease can accumulate to build up a large prevalence

Causes of increase and decrease of prevalence

Increase	Decrease
Long duration	Shorter duration
Low cure rate	High cure rate
Low case fatality	High case fatality
Increase in new cases	Decrease in new cases
Immigration of patients	Emigration of patients
Improved detection	Improved cure rate
Emigration of healthy people	Immigration of healthy people

Conclusion: Changes in prevalence may have many causes and are difficult to interpret

Uses of prevalence data

- Assessing health care needs
- Planning health services
- Measure occurrence of conditions with gradual onset
- Study chronic diseases

Incidence – (I)

- Number of new cases in a given period in a specified population
  - Time, (i.e., day, month, year) must be specified
- Measures the rapidity with which new cases are occurring in a population
- Can be expressed:
  - In absolute numbers
  - In terms of cumulated incidence
  - In terms of incidence density

Cumulated incidence - (CI)

$$CI = \frac{\text{\# of new cases}}{\text{Population at risk at the beginning}} \times 10^n$$

- Also known as:
  - Attack rate
- Assumes that the entire population at risk at the beginning was followed-up for the time period of observation

Risk

- Probability that an individual will experience a health status change over a specified follow-up period
- This assumes that the individual does not:
  - Have disease at the beginning
  - Die from other causes during follow up
- Corresponds to cumulated incidence

Incidence density - (ID)

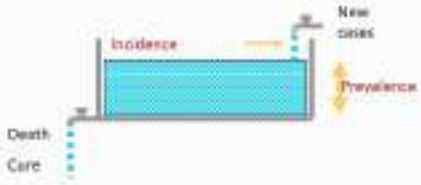
$$ID = \frac{\text{\# of new cases}}{\text{Total person-time of observation}} \times 10^n$$

- Also known as:
  - Incidence rate
- Reflects more exactly the person-time observed

Uses of incidence data

- Describe trends in diseases
- Evaluate impact of primary prevention programmes

The dynamic of incidence and prevalence



The relation between prevalence and incidence

- Prevalence depends on
  - Incidence (I)
  - Duration of the disease (D)

$$P = I \times D$$

- Change in prevalence from one time period to another may be the result of changes in incidence rates, changes in the duration of disease, or both

Patterns of incidence and prevalence

- High prevalence and low incidence
  - e.g., Diabetes Mellitus
- Low prevalence and high incidence
  - e.g., Common cold

Case fatality

- Place in relation the number of deaths from a disease to the number of cases
- Reflects severity
- Can be expressed as:
  - Proportion
  - Ratio
  - Not as rate (Although often referred to as case fatality rate)

Summary

- Prevalence is a static measure taken at a point in time
- Incidence is a dynamic measure taken over a certain time
- Mortality is calculated using population denominators to reflect burden while case fatality is calculated using cases as denominators to reflect severity

1. Which of the following must be considered while measuring occurrence of a disease?
   a) The number of people affected by the disease
   b) The population size from which the cases of disease arise
   c) The length of the time the population is followed
   d) All of the above

2. _____ is most useful for evaluating the impact of prevention programmes
   a) Point prevalence
   b) Period prevalence
   c) Case fatality
   d) Incidence

3. Which one of the following statements is true?
   a) High cure rate can increase the prevalence of a disease
   b) Low case fatality can reduce the prevalence of a disease
   c) Both 'a' and 'b' are true
   d) High cure rate and high case fatality can reduce the prevalence of a disease

4. Measures of disease frequency
   a) Incidence
   b) Prevalence
   c) Birth rate
   d) 'a' and 'b'

5. A measure that reflects severity of an acute infectious disease
   a) Case fatality ratio
   b) Incidence rate
   c) Prevalence
   d) Mortality rate

6. Incidence data can be used to measure the occurrence of disease with gradual onset
   a) True
   b) False

7. This measure reflects the impact of a disease on population in terms of death
   a) Incidence density
   b) Case fatality
   c) Disease specific mortality
   d) Attack rate

8. While measuring the frequency of a chronic disease in a community in terms of Incidence per 1000 persons per year, and point prevalence per 1000 persons, what is the expected pattern of incidence and prevalence?
   a) Low prevalence, high incidence
   b) High prevalence, low incidence
   c) Both prevalence and incidence will be similar
   d) None of the above statements are true

9. In a study among 3400 children aged 5-10 years, 16 children were diagnosed with autistic disorder. Calculate the prevalence of autism per 1000 children
   a) 4.01
   b) 5.53
   c) 3.35
   d) 4.71

10. Statistic used to estimate the risk of acquiring a disease
    a) Prevalence
    b) Incidence
    c) Mortality rate
    d) All of the above

11. What is the appropriate measure when a researcher wishes to know the burden of a particular disease in terms of the number of deaths it causes in a specified geographical region and population?
    a) Incidence density
    b) Case fatality
    c) Attack rate
    d) Disease specific mortality

12. If health policy makers want to evaluate the impact of a prevention program, which is the appropriate measure to be considered?
    a) Period prevalence
    b) Incidence
    c) Point prevalence
    d) Case fatality

13. Select the correct statement among the following
    a) Prevalence of a disease will increase when it has a high cure rate
    b) Prevalence of a disease will decrease when it has a low case fatality ratio
    c) Prevalence of a disease will increase when it has a low cure rate
    d) Prevalence of a disease will increase when it is acute in nature

14. What is the appropriate epidemiologic measure to determine the severity of an acute disease?
    a) Incidence rate
    b) Prevalence
    c) Mortality rate
    d) Case fatality ratio

15. Cumulative incidence is otherwise known as
    a) Attack rate
    b) Case fatality rate
    c) Mortality rate
    b) Morbidity rate

16. The healthcare professionals working in an intensive care unit were asked whether there has been an increase in the number of new pneumonia cases. Which of the following factor(s) is inappropriate in the calculation of cumulative incidence?

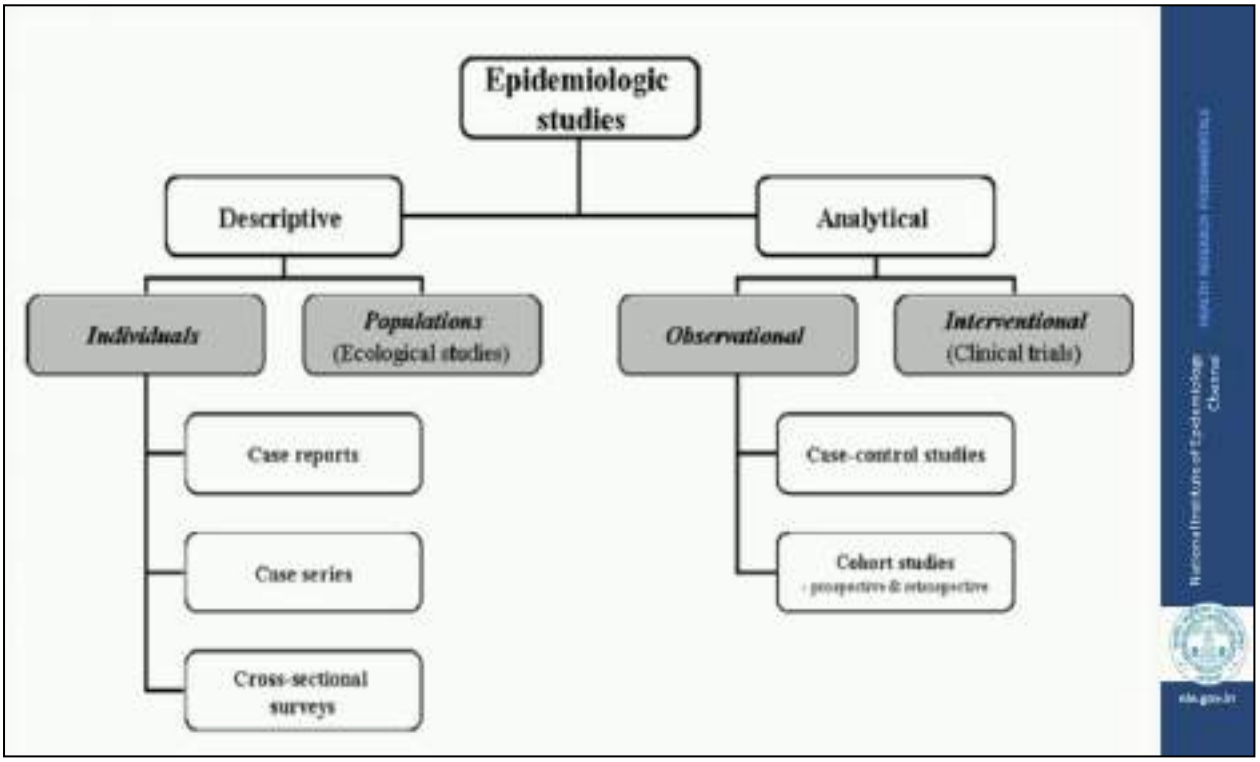
a) Number of new cases of pneumonia during a specific period of time
b) Total number of people at risk of developing the disease in that population during the same period of time
c) Pre-existing cases of pneumonia
d) Both 'a' and 'b'

17. When measuring the frequency for an acute infectious disease in a community in terms of incidence per 1000 persons per year and point prevalence per 1000 persons, how will the pattern of incidence and prevalence be?
a) High prevalence
b) Low incidence
c) Both prevalence and incidence will be similar
d) Low prevalence and high incidence

18. Among 25000 population in a city, 105 residents were identified with Hepatitis B infection. Calculate the prevalence of Hepatitis B per 1000 population.
a) 5.2
b) 4.2
c) 3.2
d) 2.2

19. Which of the following condition tends to increase the prevalence of a particular disease?
a) High cure rate
b) Low case fatality ratio
c) Short duration
d) Emigration of patients

20. Which of the following is true about incidence density?
a) Numerator has number of new cases
b) Also called cumulative incidence
c) Denominator is number of persons at risk
d) Numerator has person-years at risk



Types of descriptive studies

- Case reports
- Case series
- Ecological studies
- Cross-sectional study

Case reports

- Detailed presentation of a single case
  - New or unfamiliar diseases
  - Rare manifestations
  - Generate hypothesis regarding pathophysiological mechanism


CASE REPORT

Adenocarcinoma arising from a gastric duplication cyst with invasion to the stomach: a case report with literature review

K Kuroda, H Hatayama, T Kagawa, T Mikawa, W Yasui

Case series

- Study of larger group of patients (e.g > 10) with a particular disease
  - Larger number may allow the investigator to assess the play of chance
  - Common way of delineating the clinical pictures of a disease
- Suffers from the absence of a comparison group


Pneumocystis Pneumonia — Los Angeles

Ecological studies

- Group as the unit of analysis
- No individual-level information on the distribution of exposure and disease
- Relate whether populations with high rates of disease also have high frequency of the suspected exposure

Cross sectional surveys

- Observation of a cross-section of a population at a single point in time
  - Unit of observation and analysis: The individual
- Collect information about disease burden
  - Also known as "prevalence studies"
- Recruitment of study participants
  - Population
  - Population sample
- Observation for the presence of:
  - One or more outcomes
  - One or more exposures

Uses of cross sectional surveys

- Estimate prevalence of disease or their risk factors
- Distribution of health problem by time, place and person
  - Plan health care services delivery
- Set priorities for disease control
- Generate hypotheses
- Examine evolving trends
  - Before / after surveys
  - Iterative cross sectional surveys

Examples of research questions to be addressed through surveys

- What is the prevalence of hypertension in a city?
- How satisfied are patients attending government hospitals in Chennai?
- What is the prevalence of physical inactivity among school children?

Advantages and limitations of cross sectional survey

- Advantages
  - Fairly quick and easy to perform
  - Less expensive
- Limitations
  - Not useful to study disease etiology
  - Not suitable for the study of rare diseases

Cross sectional survey: major limitation

- Prevalent cases (Old and new cases)
- Exposure and outcome examined at the same time. e.g.
  - Obesity and diabetes

Take home messages

- Case reports and case series are useful for uncommon clinical manifestations
- Ecological studies can be used to relate group level data and generate hypothesis
- Cross sectional surveys help to measure the burden or magnitude of health condition

1. Study design(s) useful for describing uncommon clinical manifestations
   a) Case reports
   b) Case series
   c) Both 'a' and 'b'
   d) Ecological study

2. Cross-sectional studies are used to
   a) Estimate prevalence
   b) Generate hypotheses
   c) Describe trends
   d) All of the above

3. In a cross-sectional study, we can observe one or more outcomes
   a) True
   b) False

4. Which one of the following is useful to measure the burden or magnitude of a disease or risk factor?
   a) Case-control study
   b) Cross-sectional study
   c) Case report
   d) Case series

5. Which of the following is wrong about descriptive study designs?
   a) Describe the study outcome for 1 group
   b) Compare the study outcomes for 2 group
   c) Calculate the incidence for surveillance data
   d) Calculate prevalence for cross-sectional study

6. Descriptive epidemiology study designs can answer all of the following questions EXCEPT:
   a) Who?
   b) When?
   c) Where?
   d) Why?

7. Which one of the following study designs does not employ comparison groups to answer the primary study objectives?
   a) Cross-sectional study
   b) Cohort study
   c) Ecological study
   d) Clinical trials

8. Unit of observation in the cross-sectional study is
   a) Individual

b) Group
c) Both 'a' and 'b'
d) None of the above

9. Case reports can include presentation of
a) Unique features/symptoms of a disease
b) Rare manifestation of common disease
c) New or unfamiliar diseases
d) All of the above

10. Advantage of the ecological study is
a) Relate rate of disease and exposure
b) Useful to test hypothesis
c) Useful to study rare diseases
d) All of the above

11. A researcher can assess the following by conducting a descriptive study EXCEPT
a) Population in which the disease was prevalent
b) Period in which the disease occurred
c) Risk factors of the disease
d) Place distribution of the disease

12. The following study design provides group exposure and group response/outcome without knowing the individual exposure and response for a specific health problem
a) Ecological study
b) Cross sectional survey
c) Case report
d) Case series

13. Which of the following study design will be helpful if the department of health wants to know the burden of a particular disease?
a) Ecological study
b) Cross sectional survey
c) Case series
d) Case report

14. A clinician comes across an unusual presentation of a particular neurological disorder. If the clinician describes this single case in detail and publishes the same in a journal, then it will be called
a) Analytical study
b) Case report
c) Cross sectional survey
d) Ecological study

15. The advantage of an ecological study is that
a) It is analytical in nature
b) It will cover individual level information on risk factors and disease
c) It will be useful to test hypotheses
d) It will be useful to generate hypotheses

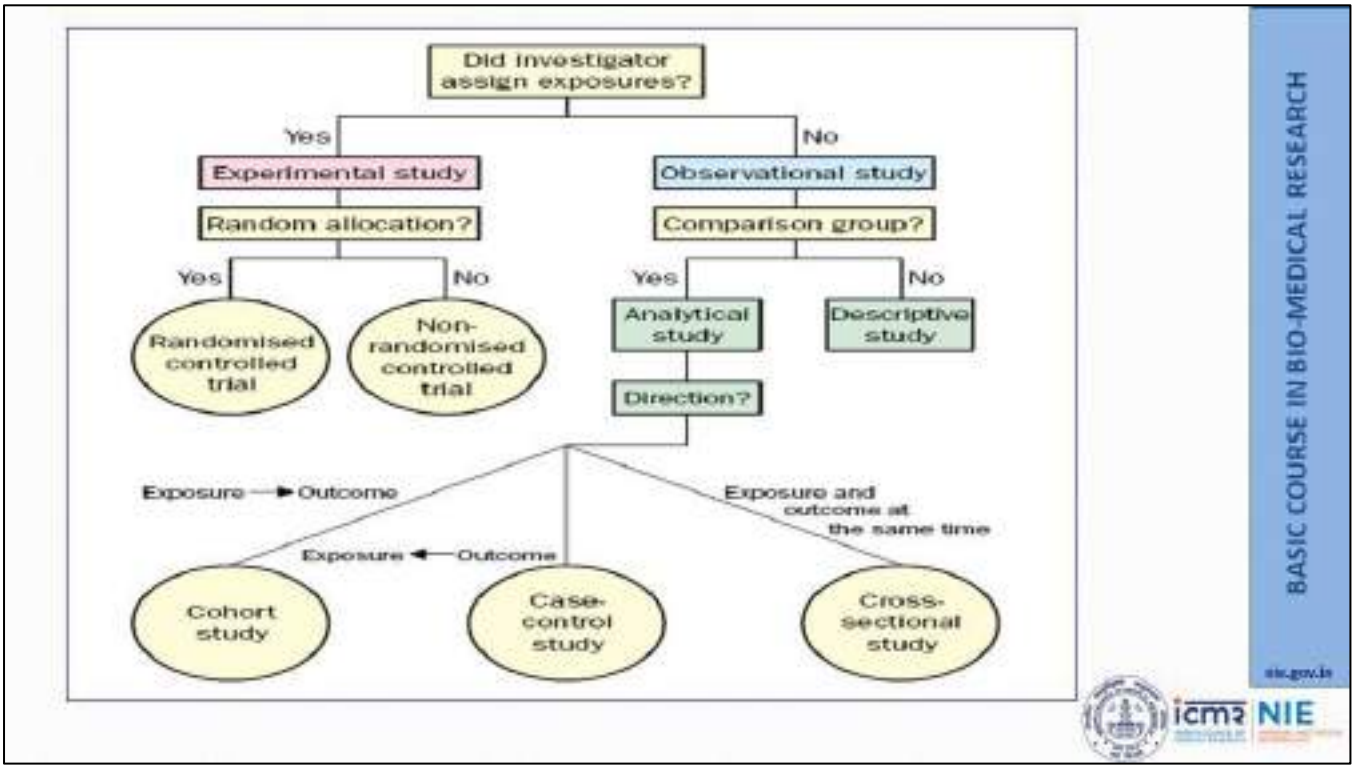
16. In a tertiary care hospital, a surgeon collected information on quality of life and outcome among a small group of (about 15) post-operative patients after using a novel surgical device. But this is not sufficient to establish the efficacy of the surgical device because
a) There is no comparison group
b) There is no information of risk factors
c) We do not have details of the outcome
d) We do not have individual level data

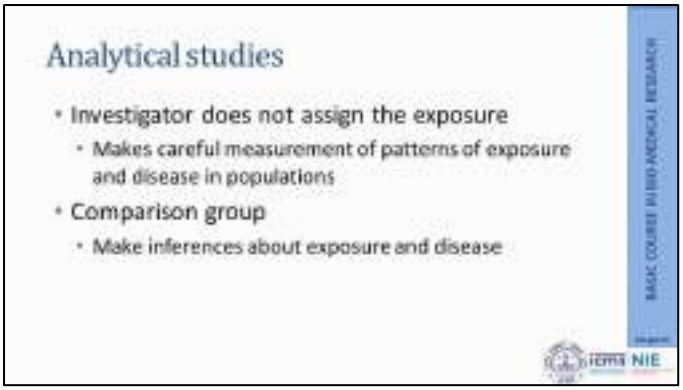
17. Population census is a
a) Cross sectional survey
b) Ecological study
c) Analytical study
d) None of the above

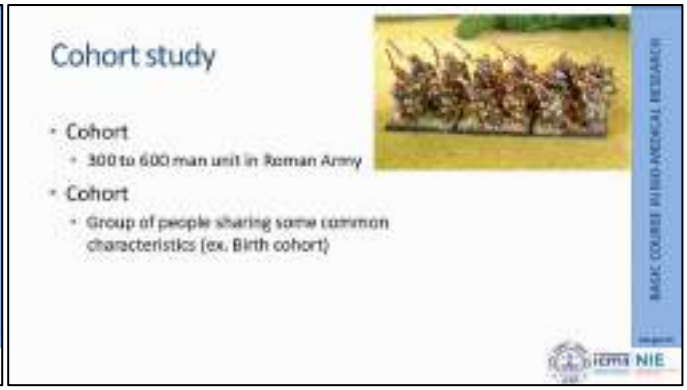
18. One of the major limitations of a cross-sectional study is that
a) It is time consuming
b) It has lower validity
c) It does not establish disease etiology
d) It requires a large sample size

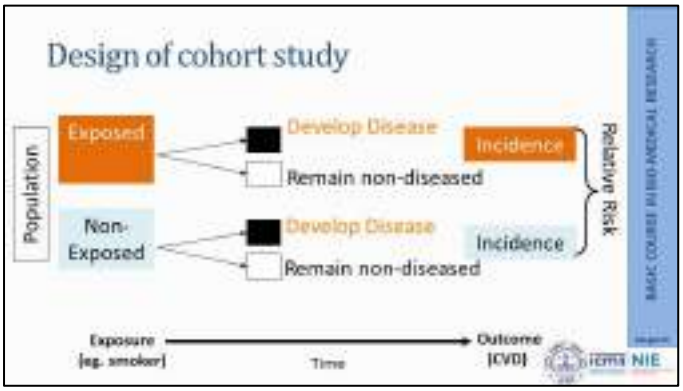
19. Characteristic of a cross sectional study is that
a) We can calculate the incidence of a disease
b) We can test a hypotheses
c) It is difficult to conduct
d) Exposure and outcome are assessed at the same time

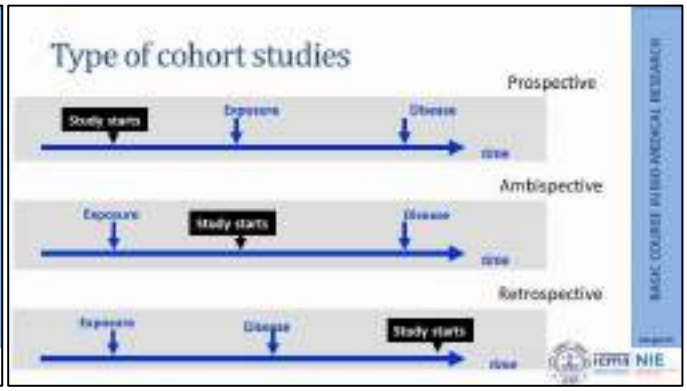
20. If a researcher wishes to estimate the incidence of Myocardial infarction cases among a group of women using oral contraceptive pills followed up for 10 year, the researcher has to carry out
a) Case series
b) Cohort study
c) Cross sectional study
d) Ecological study

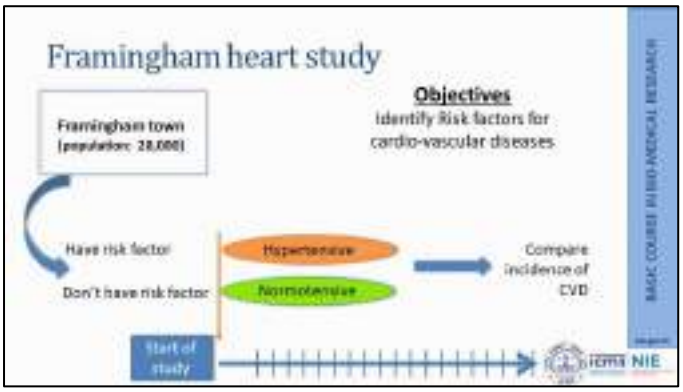

Did investigator assign exposures?
Yes
No
Experimental study
Observational study
Random allocation?
Comparison group?
Yes
No
Yes
No
Randomised controlled trial
Non-randomised controlled trial
Analytical study
Descriptive study
Direction?
Exposure → Outcome
Exposure ← Outcome
Exposure and outcome at the same time
Cohort study
Case-control study
Cross-sectional study
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE

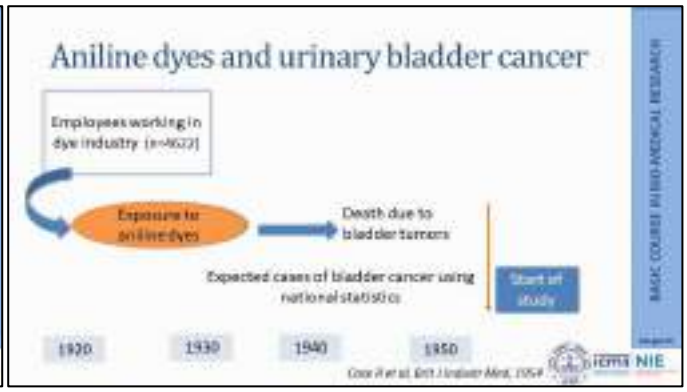

Analytical studies
Investigator does not assign the exposure
Makes careful measurement of patterns of exposure and disease in populations
Comparison group
Make inferences about exposure and disease
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE


Cohort study
Cohort
300 to 600 man unit in Roman Army
Cohort
Group of people sharing some common characteristics (ex. Birth cohort)
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE


Design of cohort study
Population
Exposed
Non-Exposed
Develop Disease
Remain non-diseased
Develop Disease
Remain non-diseased
Incidence
Incidence
Relative Risk
Exposure (eg. smoker)
Time
Outcome (CVD)
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE


Type of cohort studies
Prospective
Study starts
Exposure
Disease
Time
Ambispective
Exposure
Study starts
Disease
Time
Retrospective
Exposure
Disease
Study starts
Time
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE


Framingham heart study
Framingham town (population: 28,000)
Objectives
Identify Risk factors for cardio-vascular diseases
Have risk factor
Don't have risk factor
Hypertensive
Normotensive
Compare incidence of CVD
Start of study
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE


Aniline dyes and urinary bladder cancer
Employees working in dye industry (n=4622)
Exposure to aniline dyes
Death due to bladder tumors
Expected cases of bladder cancer using national statistics
Start of study
1920
1930
1940
1950
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE

Elements of cohort study

1. Selection of study populations
2. Gathering baseline information
3. Follow-up
4. Analysis

Selection of study population

- General population cohorts or a sub-set
  - Framingham heart study
  - Nurses health study
- Special exposure cohorts
  - Occupational groups

Gathering baseline information

- Objective
  - Valid assessment of exposure status of members of cohort
    - Identification data
    - Exclude individuals having disease at baseline
    - Define individuals at risk
    - Obtain data on co-variables (other exposure variables)

Choice of comparison group

- Internal comparison group
  - Unexposed persons in the population
- External comparison group
  - When internal comparison group not available
  - Ex: Observed number of bladder cancer deaths in aniline dye industry compared with expected cases

Follow-up

- Objectives
  - Uniform and complete follow-up of all cohort members
    - Uniform surveillance in exposed and unexposed groups
  - Complete ascertainment of exposures and outcome/s
  - Standardized diagnosis of outcome events

Presentation of the data in a cohort study in a 2 x 2 table

	Diseased	Non-diseased	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
	a+c	b+d	a+b+c+d

Known at the start of the study

Relative risk

	Diseased	Non-diseased	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
	a+c	b+d	a+b+c+d

Incidence of disease in exposed = a/a+b

Incidence of disease in unexposed = c/c+d

Interpreting Relative risk

- RR=1
  - Incidence in exposed and unexposed is same
  - Exposure is not associated with disease
- RR > 1
  - Incidence in exposed is higher than unexposed
  - Exposure is positively associated with disease
- RR < 1
  - Incidence in exposed is lower than unexposed
  - Exposure is negatively associated with disease

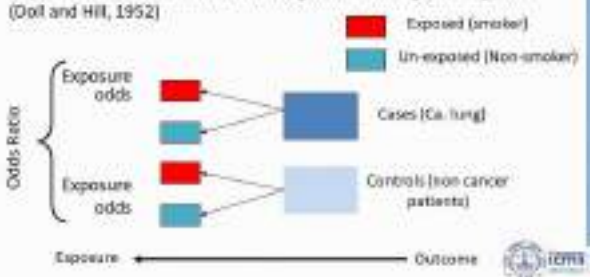
Cohort study – Strengths and weaknesses

- Strengths
  - Allows calculation of incidence
  - Examine multiple outcomes for a given exposure
  - Clarity of temporal sequence
  - Good for investigating rare exposures
- Weakness
  - May have to follow large numbers of subjects for a long time.
    - Expensive and time consuming.
  - Not good for rare diseases.
  - Not good for diseases with a long latency.
  - Differential loss to follow up can introduce bias.

Case control study

Design of case-control study

Objective: Test association between cigarette smoking and lung cancer (Doll and Hill, 1952)



Elements of case control study

1. Selection of cases
2. Selection of controls
3. Information on exposure
4. Analysis

Selection of cases

- All people in source population who develop the disease of interest
  - Sample of cases
    - Independent of the exposure under study
- Clear definition of outcome studied
- Prevalent vs. incident cases
  - Prevalent cases may be related more to survival with disease than to development of disease

Sources of cases

- Hospital/clinic based cases
  - Easier to find
  - May represent severe cases
- Population based (cancer registry)
  - not biased by factors drawing a patient to a particular hospital

Selection of controls

- Represent the distribution of exposure in the source population of cases
  - Selected from the same source population that gives rise to the cases
- Selected independently of their exposure status

Selection of controls

- Population based
  - Sampling of the general population
- Health care facility based
  - Patients with other diseases
- Case-based
  - Friends, Neighbourhood

Collecting good data on exposure

- Objectively
  - Reproducibility of exposure measurement
- Accurately
  - Information reflecting as closely as possible the effect of exposure
- Precisely
  - Quality management in exposure measurement

Presentation of the data of a case-control study in a 2 x 2 table

	Cases	Controls	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
	a+c	b+d	a+b+c+d

Known at the start of the study

Odds ratio

	Cases	Controls	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
	a+c	b+d	a+b+c+d

Odds that case was exposed = Probability that case was exposed / Probability that case was not exposed = a/c

Odds that control was exposed = Probability that control was exposed / Probability that control was unexposed = b/d

Odds ratio=(a/c)/(b/d) = ad/bc

Interpreting Odds Ratio

- OR=1
  - Odds of exposure among cases and controls are same
  - Exposure is not associated with disease
- OR > 1
  - Odds of exposure among cases are higher than controls
  - Exposure is positively associated with disease
- OR < 1
  - Odds of exposure among cases are lower than controls
  - Exposure is negatively associated with disease

1. Exposure is not assigned by the investigator in the following study design
   a) Cohort
   b) Case-control
   c) Cross-sectional
   d) All of the above

2. Which of the following statement(s) is true about the cohort study?
   a) It is not suitable for disease with a long latency period
   b) Loss to follow up can introduce bias
   c) Both 'a' and 'b'
   d) None of the above

3. Which of the following study design is better suited to demonstrate a temporal association between exposure and disease?
   a) Cross-sectional study
   b) Case-control study
   c) Cohort study
   d) Ecological study

4. Relative risk of more than 1 indicates
   a) Incidence in unexposed is higher than exposed
   b) Incidence in exposed and unexposed are same
   c) Incidence in exposed is higher than unexposed
   d) Relative risk is significant

5. If the odds of exposure among cases is lower than the odds of exposure among the controls, the odds ratio will be
   a) More than 1
   b) Less than 1
   c) It depends on other factors
   d) None of the above

6. All babies born in a particular year will form a birth cohort
   a) True
   b) False

7. Cohort study is suitable for rare diseases
   a) True
   b) False

8. Which of the following is NOT true regarding case-control study?
   i. Appropriate for study of rare outcome
   ii. More time consuming than cohort study
   iii. Multiple exposures can be examined
   iv. Relatively expensive compared to cohort study
   a) Both (i) and (ii)
   b) Both (ii) and (iii)
   c) Both (ii) and (iv)
   d) Both (iii) and (iv)

9. What is an appropriate measure of statistical association in a cohort study?
   a) Prevalence ratio
   b) Risk ratio
   c) Odds ratio
   d) Pearson's correlation coefficient

10. The entire population of a given community is screened and all those judged as being free of Colon cancer are questioned extensively about their diet. These people are then followed-up for several years to see whether their eating habits will predict their risk of developing Colon cancer - This is an example of
    a) Case-control study
    b) Clinical trial
    c) Cross-sectional study
    d) Cohort study

11. Exposure is assigned by the investigator in which of the following epidemiological study?
    a) Case-control
    b) Cross-sectional
    c) Experimental
    d) Cohort

12. When a group of people with defined characteristics are followed up to determine incidence is known as
    a) Case series
    b) Cohort
    c) Case control
    d) Experimental

13. Relative risk is a
    a) Rate
    b) Ratio
    c) Proportion
    d) None of the above

14. Relative risk of one in a cohort study indicates
    a) Incidence in unexposed is higher than exposed
    b) Incidence in exposed is higher than unexposed
    c) Relative risk is significant
    d) Incidence in the exposed and unexposed groups are same

15. Women aged above 35 years were screened for the HPV (Human papilloma virus) infection and those who had HPV infection were then followed for several years to predict the risk of developing cervical cancer. This study is known as
    a) Prospective cohort
    b) Retrospective cohort
    c) Case control
    d) Cross sectional

16. Which of the following is appropriate regarding a cohort study?
    i. Multiple exposures can be examined
    ii. Appropriate for studying rare exposures
    iii. Expensive and time consuming
    iv. Appropriate for studying rare diseases
    a) Both (i) and (ii)
    b) Both (iii) and (iv)
    c) Both (ii) and (iv)
    d) Both (ii) and (iii)

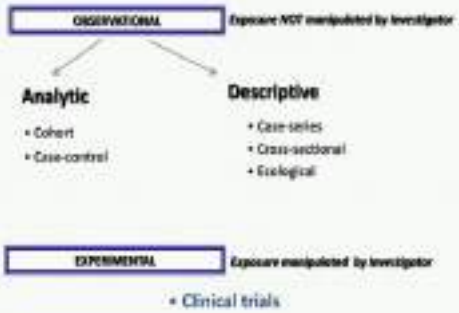
17. Which of the following statement regarding the cohort study is FALSE?
    a) Suitable to study a disease with long latency period
    b) Loss to follow up can introduce bias
    c) Relative risk can be calculated
    d) Temporal association with the risk factor can be established

18. Odds ratio of more than one indicates
    a) Odds of exposure among cases is lower than the odds of exposure among the controls
    b) Odds of exposure among cases is equal to the odds of exposure among the controls
    c) Odds of exposure among cases is higher than the odds of exposure among the controls
    d) Exposure is negatively associated with the disease

19. If there is a comparison group in an epidemiological study design, it is called
    a) Descriptive
    b) Analytical
    c) Ecological
    d) None of the above

20. Which of the following statements about case control/cohort studies is correct?
    a) Case control study always establishes temporal association
    b) Cohort study establishes temporal association
    c) Cohort has lower level of evidence than case-control
    d) Do case control for rare exposures and cohort for rare diseases

Study Designs



• Clinical trials

Randomized Controlled Trials

One of the main scientific advances in methods of clinical research in the 20th century. They are considered as the methodologic standard of excellence and gold standard for scientific experiments.

Significance of clinical trials

- Clinical trials translate results of basic scientific research into better ways to prevent, diagnose, or treat disease
- Research studies involving people are designed to answer scientific questions and find better ways to prevent, diagnose, or treat disease

Randomized controlled clinical trials

- A clinical trial is a planned experiment designed to assess the efficacy of prophylactic / diagnostic / therapeutic agents, devices, regimens, procedures etc. applied to human subjects
- It essentially involves comparing the outcomes in a group of patients treated with a test treatment with those observed in a comparable group of patients receiving a control treatment where patients in both groups are enrolled in a prospective study, treated or exposed to intervention and followed over the same period

The objectives of clinical trials

Clinical trials are normally conducted to evaluate new forms of therapy or prevention methods such as

- New drugs/ treatment
- New medical / health care technology
- New organization/ delivery system of health care
- New methods of primary prevention
- New programs of screening or early detection

Schematics of a therapeutic clinical trial

Randomization

Randomization ensures that participants have an equal chance to be assigned to one of two or more groups:
A. One group gets the most widely accepted treatment (standard treatment/ gold standard)
B. The other gets the new treatment being tested, which researchers hope and have reason to believe will be better than the standard treatment

Randomization provides the best way to prove the effectiveness of a new agent or intervention by ensuring that
A. All groups are as similar as possible
B. Confounding, co-interventions and bias in outcome ascertainment is minimized

Blinding in clinical trials helps in balancing groups during follow-up

- Blinding can be at the level of
  - participants (single blinding)
  - Participants and investigators (double blinding) and
  - Participants, investigators and analysts (triple blinding)
- Blinding helps to eliminate
  - Co-intervention: participants use other therapy or change behavior or study staff, medical providers, family or friends treat participants differently
  - Biased outcome ascertainment: participants may report symptoms or outcomes differently or physicians or investigators may elicit symptoms or outcomes differently

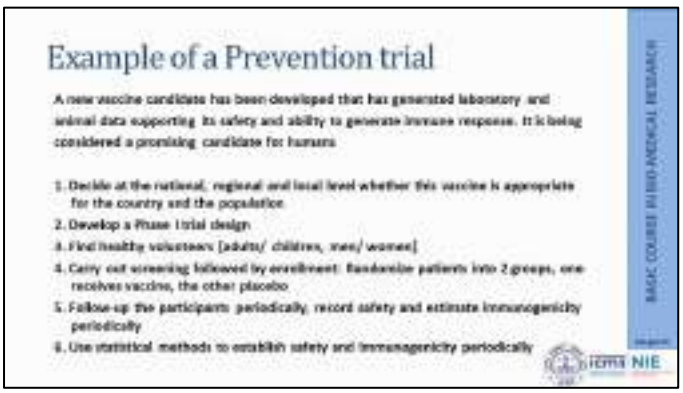
Phases in clinical trials and objectives

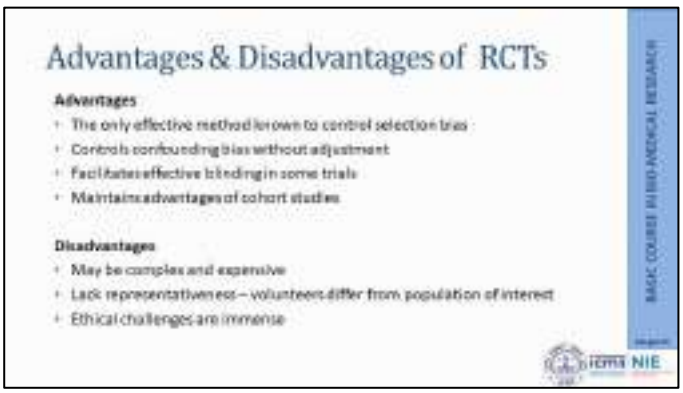
Trial phase	End-points/ objectives	Sample size and participants
Phase I	Safety Acceptability	Up to 50 Healthy volunteers
Phase II	Long-term safety Dose and schedule Early indications of efficacy	100 to 500 Low risk
Phase III	Effectiveness	1000 and more High risk
Phase IV	Post-marketing surveillance	1000 and more Community based

Example of a therapeutic trial

To study if a new drug regimen [NDR] effectively lowers viral load and improves CD4 counts in HIV infected persons compared to standard therapy [HAART]

1. Identify HIV infected persons, define inclusion and exclusion criteria
2. Randomize patients into 2 groups, one receives NDR, the other HAART
3. Follow-up periodically, estimate viral load and CD4 counts periodically
4. Use statistical methods to see if there are differences between viral load and CD4 counts in the two groups





1. One of the cornerstones of the randomized controlled trials is
   a) Recruitment
   b) Randomization
   c) Blinding
   d) Placebo

2. Randomized clinical trials can be best described as
   a) Experimental studies
   b) Analytic studies
   c) Descriptive studies
   d) Observational studies

3. Which of the following is/are true in a clinical trial?
   a) Sample size determination
   b) Approval from regulatory authority
   c) Agreement between the investigators and sponsors
   d) All of the above

4. A pharmacologically inactive agent that investigators administer to participants in the control group of a trial
   a) Comparator drug
   b) Placebo
   c) Conjugate
   d) Drug under investigation

5. Key methodological components of a Randomized Controlled Trials are
   a) Use of a control to which the experimental intervention is compared
   b) Random assignment of participants to intervention
   c) Taking informed consent from all study participants
   d) All of the above

6. Double-blinding in a clinical trial involves
   a) Participants before and after study
   b) Participants and investigators
   c) Investigators and analysts
   d) Participants and analysis

7. The purpose of a double-blinding in a clinical trial is to
   a) Achieve comparability of all arms of a clinical trial
   b) Avoid observer and participant bias
   c) Avoid observer bias and sampling variation
   d) Avoid subject bias and sampling variation

8. What is the purpose of randomization in a clinical trial?
   a) Get better power for data analysis
   b) Generalizing the study findings to the population which is not studied
   c) Achieve balance in baseline characteristics
   d) Guarantee that the statistical tests have valid significance levels

9. Which phase of a clinical trial is referred to as post-marketing surveillance?
   a) Phase 1
   b) Phase 2
   c) Phase 3
   d) Phase 4

10. Long-term adverse effects and efficacy of a new drug can be tested in which of the following phases of a clinical trial?
    a) Phase 1
    b) Phase 2
    c) Phase 3
    d) Phase 4

11. Which of the following is incorrect in case of a clinical trial?
    a) All clinical trials must be blinded
    b) Randomization is a critically important step in a clinical trial
    c) All clinical trials must be approved by Institutional Ethics Committee before initiation
    d) It is mandatory to register clinical trials with Clinical Trials Registry of India

12. Which of the following procedures ensure safety of the clinical trial participants?
    a) Adverse events reporting
    b) Serious adverse events reporting
    c) Periodic follow-up
    d) Review by Data Safety Monitoring Board
    e) 'a', 'b', 'c', and 'd'
    f) 'a', 'b' and 'd'

13. All the following correctly describe a clinical trial, EXCEPT-
    a) It has all advantages of a cohort study
    b) It is possible to analyse the confounders

c) Loss to follow up of study participants does not affect the study outcome

d) Appropriate implemented informed consent procedure as well as long-term care and support to trial participants help to overcome several ethical concerns

14. Which of the following can be considered as an advantage of a double blinding in a randomized controlled trial?

a) Equally distributes known and unknown confounders in experiment and control arm

b) Ensures that participants adhere to the protocol

c) Gives benefits of an intervention to some of the study participants

d) Prevent bias that arises from researchers being able to influence the data due to knowledge of allocated groups

15. Biased outcome ascertainment results from:

a) Participants reporting symptoms or outcomes differently

b) Investigators eliciting symptoms or outcomes following a standardized technique

c) None of the above

d) Both "a" and "b"

16. In a clinical trial, what is the main purpose of randomization?

a) To get more power for data analysis

b) To reduce investigator bias

c) To get groups with comparable baseline characteristics

d) To ensure optimum number of participants in each trial arm

17. Which of the following can eliminate the problem of Co-intervention?

a) Random sampling

b) Allocation concealment

c) Informed consent

d) Blinding

18. Which of the following is not true in case of a clinical trial?

a) Clinical trials are planned experiments designed to assess the efficacy of an intervention

b) Clinical trials usually involve comparing the outcomes in two or more groups of individuals

c) Clinical trials are usually free from selection bias

d) Clinical trials are usually prospective in nature

19. Which of the following trials assesses effectiveness of a new vaccine?

a) Phase 1 trial done in healthy volunteers

b) Phase 2 trial done in a susceptible population

c) Phase 3 trial done in healthy volunteers

d) Phase 3 trial done in a susceptible population

20. Which of the following is not true about a randomized control trial?

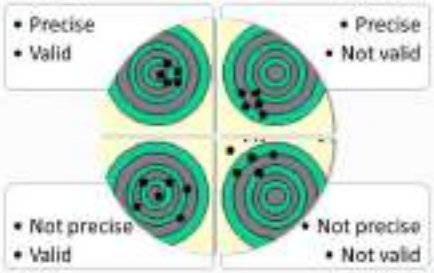
a) Baseline characteristics of intervention and control groups must be similar

b) Investigator bias can be minimized by double blinding

c) The sample size depends on the hypothesis being tested

d) Drop outs should be excluded from the analysis

Accuracy = Precision and Validity



Random errors
CHANCE
Variability in estimation due to unknown/uncontrolable causes
- Sampling error
- Measurement error

Errors in estimation

Systematic errors
BIAS
A process that tends to produce results that depart systematically from true values

Association between exposure and outcome among the study participants differs from the causal association in the population

Threats to validity in epidemiologic studies

Selection bias → Information bias → Confounding

Threats to validity in epidemiologic studies

Selection bias

Do study participants accurately represent target population?

Bias results from procedures used to select individuals for inclusion in the study and/or analyses

Selection bias in epidemiological studies

- **Surveillance** - Systematic notification of exposed cases
- **Screening / diagnosis** - Systematic case search among exposed
- **Admission to health care facilities** - Systematic admission of:
  - Case patients exposed/ unexposed
  - Control subjects exposed/ unexposed
- **Selective survival** - Systematic inclusion of cases who survived and who may be more or less exposed
- **Non response / loss to follow up** - Systematic inclusion of subjects more likely to participate who may be:
  - More or less exposed
  - More or less at risk

Dealing with selection bias

Designing stage of a study

- Use incident cases, not prevalent cases
- Case control studies
  - Use population-based design
  - Apply same eligibility criteria for selecting cases and controls
  - Both cases and controls undergo the same diagnostic procedures and intensity of disease surveillance

Dealing with selection bias

Data collection stage of the study

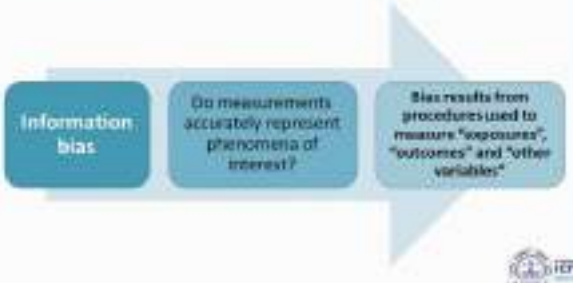
- Minimize nonresponse, nonparticipation and loss to follow-up (Cohort studies)
- Keep a record of all losses and collect baseline data on them
- Make sure that diagnosis of disease is not affected by exposure status (blinding)

Dealing with selection bias

Analysis stage of study

- Compare non-responders/dropouts with responders/non-dropouts with respect to baseline variables
  - Large differences strongly suggest selection bias
  - Small differences do not rule out selection bias
- Use study results and external information to deduce the direction of biases and assess magnitude of biases
  - Do sensitivity analysis

Threats to validity in epidemiologic studies



Information bias in epidemiological studies

- **Case control study**
  - Collection of information leaning towards specific exposure status
  - Recall - Cases may recall exposure more than controls
  - Better exposure data available on cases compared to controls
- **Cohort study**
  - Collection of information leaning towards specific outcome status
  - Better outcome data available on exposed compared to unexposed
- **Investigator** - Systematic collection of information supporting expected conclusions (Unconsciously or Consciously)
  - May be examined in the analysis
- **Prevarication** - Systematic distortion of the truth by subjects

Dealing with information bias

- Precise operational definitions of variables
- Detailed measurement protocols
- Repeated measurements on key variables
- Training, certification and re-certification
- Data audits (of interviewers, of data centers)
- Data cleaning – visual, computer
- Re-running all analyses prior to publication

BASIC COURSE IN BIO-MEDICAL RESEARCH

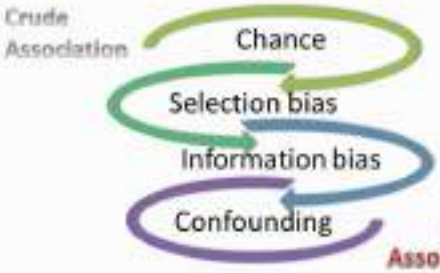
Threats to validity in epidemiologic studies

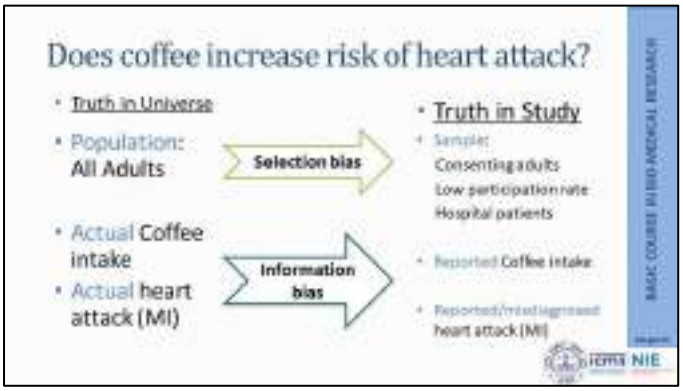
Confounding in epidemiological studies

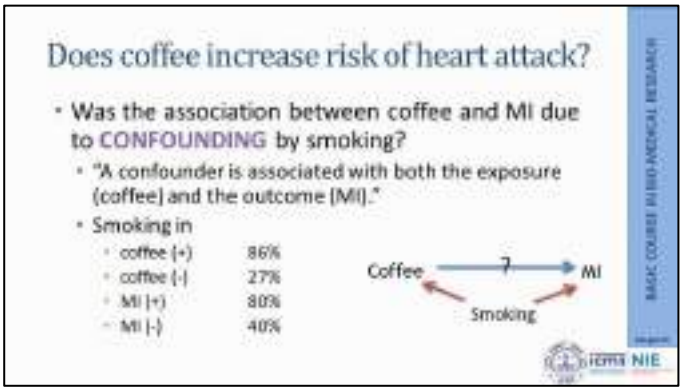
Dealing with confounding

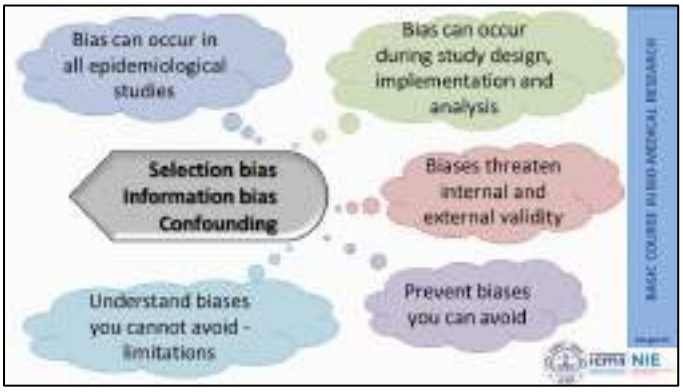
- Design stage
  - Restriction
  - Matching
  - Randomization
    - Experimental studies
- Analysis stage
  - Stratification
  - Multivariate analysis

How to evaluate associations?









1. Obtaining an estimate that is generalizable to relevant study population in a research study is
   a) External validity
   b) Internal validity
   c) Bias
   d) Confounding

2. Any process that tends to produce results that depart systematically from true values in a research study
   a) Chance
   b) Bias
   c) Random error
   d) Effect Modification

3. Systematic selection of more number of exposed participants with the higher risk of outcome in a cohort study will result in
   a) Selection bias
   b) Information bias
   c) Confounding
   d) Random error

4. The effect of the exposure of interest on the outcome is distorted because of the effect of extraneous factors that are related to both the exposure and outcome. This phenomenon is called
   a) Correlation effect
   b) Confounding
   c) Recall bias
   d) Measurement error

5. Biases can occur during which stage of research study?
   a) Study design
   b) Study implementation
   c) Data analysis
   d) At any of the above stages

6. All are true regarding measures to reduce information bias, EXCEPT
   a) Precise operational definitions of all variables
   b) Detailed measurement protocols
   c) Adequate sample size
   d) Training, Certification and re-certification of data collectors

7. Variability in estimation due to unknown/uncontrollable factors
   a) Chance
   b) Bias
   c) Confounding
   d) Effect modification

8. All are true regarding confounding in an epidemiological study, EXCEPT
   a) May simulate an association that does not exist
   b) May increase or decrease the strength of association
   c) May not reveal an association that does exist
   d) Always change the direction of effect

9. The method which can used to alleviate confounding during data analysis in an epidemiological study
   a) Multivariate analysis
   b) Restriction
   c) Matching
   d) Randomization

10. To reduce selection bias in case-control studies, all of the following are true EXCEPT
    a) Use population based design
    b) Apply different eligibility criteria for selecting cases and controls
    c) Both cases and controls undergo the same diagnostic procedures
    d) Avoid hospital based design

11. Obtaining an accurate estimate of disease frequency and effect of exposure on health outcomes in study population pertains to
    a) External Validity
    b) Internal Validity
    c) Bias
    d) Confounding

12. Blinding in an epidemiological study is a way to deal with
    a) Chance
    b) Selection Bias
    c) Information Bias
    d) Sampling Error

13. Better recall of exposure only among the cases in a case control study can result in
    a) Information bias
    b) Confounding
    c) Investigator bias
    d) Selection bias

14. The ability of a tool to correctly measure what it is supposed to measure is called as
    a) Precision
    b) Validity
    c) Reliability
    d) Consistency

15. Bias may distort the association between exposure and outcome among the study participants
    a) True
    b) False

16. A case control study was conducted to know the effect of smoking on lung cancer among hospitalized patients. The controls were recruited from patients admitted to the respiratory ward for other conditions. What type of bias will be introduced by virtue of recruiting controls from the hospital who are potentially different from the general population?
    a) Selection bias
    b) Information bias
    c) Confounding
    d) Random error

17. A researcher studied the effect of coffee drinking on Myocardial Infarction. The effect of coffee drinking on Myocardial Infarction was distorted because of the presence of a third factor, i.e. smoking. This phenomenon is called as
    a) Correlation effect
    b) Confounding
    c) Recall bias
    d) Measurement error

18. Which of the following method is used to address for known confounders at the designing stage of a study?
    a) Matching
    b) Regression
    c) Stratification
    d) Adjusted analysis

19. Systematic distortion of the truth by study subjects is called as
    a) Plagiarism
    b) Chance
    c) Confounding
    d) Prevarication

20. Crude association in the presence of a confounder is the actual causal association
    a) True
    b) False

Quantitative versus Qualitative research methods

	Quantitative	Qualitative
Data	Numbers	Text
View of the world	Social reality - objective, measurable, external to individual ETIC	Social reality – subjectively interpreted and experienced EMIC
Logic of enquiry	Deductive – testing formal hypotheses	Inductive – understanding of social processes derived from data
Research design	Ensures repeatability	Interpretation of responses
Validity	Objective (reliability)	Subjective (credibility)
Cross-cultural generalizations	Application of the same observation method to different cultures	Require conversion in abstract inter-cultural categories

BASIC COURSE IN BIO-MEDICAL RESEARCH

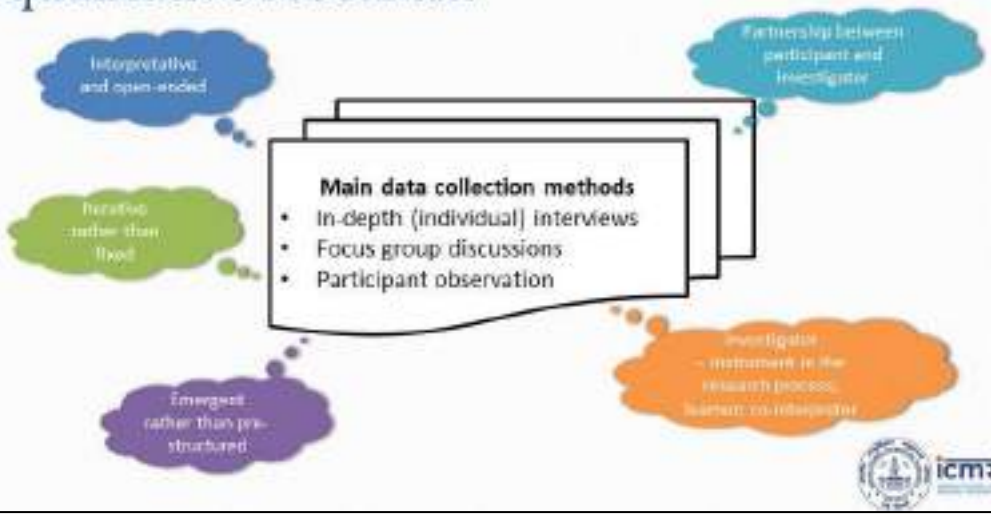
icmr NIE

When to use qualitative research methods?

BASIC COURSE IN BIO-MEDICAL RESEARCH

icmr NIE

What are the methods used in qualitative research?



BASIC COURSE IN BIO-MEDICAL RESEARCH

icmr NIE

In-depth (Individual) Interviews

- Open-ended interviews
  - Discover the interviewee's own framework of meanings
  - Obtains rich, contextualized, in-depth information
  - Avoid imposing the researcher's structures and assumptions
- Technique
  - Follows interview guide
  - Probes
  - Reflecting on remarks made by the informant
  - Collects respondent's perspective and words
  - Level of structure varies
- When to use?
  - Complex subject matter and knowledgeable respondents
  - Highly sensitive subject matter
  - Geographically dispersed respondents
  - Peer pressure an issue - social desirability a threat

In-depth (Individual) Interviews

Advantages

- Most in-depth - Why behaviors are practiced?
- Data on how people think and talk; conceptualizations of behavior
- Exact words & language people use amongst themselves
- "Emic" perspective = insider's perspective

Disadvantages

- Based on a few people, usually not systematic sample, but purposeful or convenience sample
- Not generalizable
- Interviews very long, lots of data! Time consuming to analyze
- Researchers opinions of what the data means

Focus Group Discussions

- Open-ended group interviews that promote discussions between participants on specific topics
- Usually 6-8 'similar' participants
  - Similar age, gender, socio-economic status, education, others...
    - Similar cognitive structure
    - Similar perceptions of their social environment
    - Similar normative beliefs
- Moderator and note-taker
- Flexible interview guide
- When to use?
  - Group interaction important
  - Cost and timing issues
  - Idea generation
  - Problem identification and definition goal
  - Identify local/group specific vocabulary/terminology
  - Evaluating messages for an intervention

Focus Group Discussions

Advantages

- Some people more comfortable and talk more openly in group settings
- Natural way for some people to talk about problems and personal issues in some cultures
- Collects information on social norms
- Can provide lots of data in a limited amount of time

Disadvantages

- Difficult to access practice of very personal or sensitive behaviors in groups
- Not GENERALIZABLE
  - Subject to dominant personalities
  - Sensitive to biased analysis
- Transcribing time consuming - often 30-40 pages each
  - Difficult to identify speakers
- Analytic challenge!

Participant Observation

- The researcher becomes participant in social event or group under study and records observations
- Advantages
  - Data is very deep and detailed
- Disadvantages
  - Difficult to systematically collect; especially in middle of important moment - hard to take notes so details may be forgotten
  - Analytic methods for observation notes not well described

Qualitative data (text) analysis

Grounded theory

- Transcripts of interviews
- Potential analytic categories— themes
- Coding text into categories
- Relations among categories
- Build theoretical models
- Exemplars - quotes from interviews

Content analysis

- Theoretical framework
- Set of codes for variables in the theory
- Applying codes systematically to a set of texts
- Unit-of-analysis-by-variable matrix from the texts and codes
- Statistical analysis of matrix

How to use qualitative research methods?

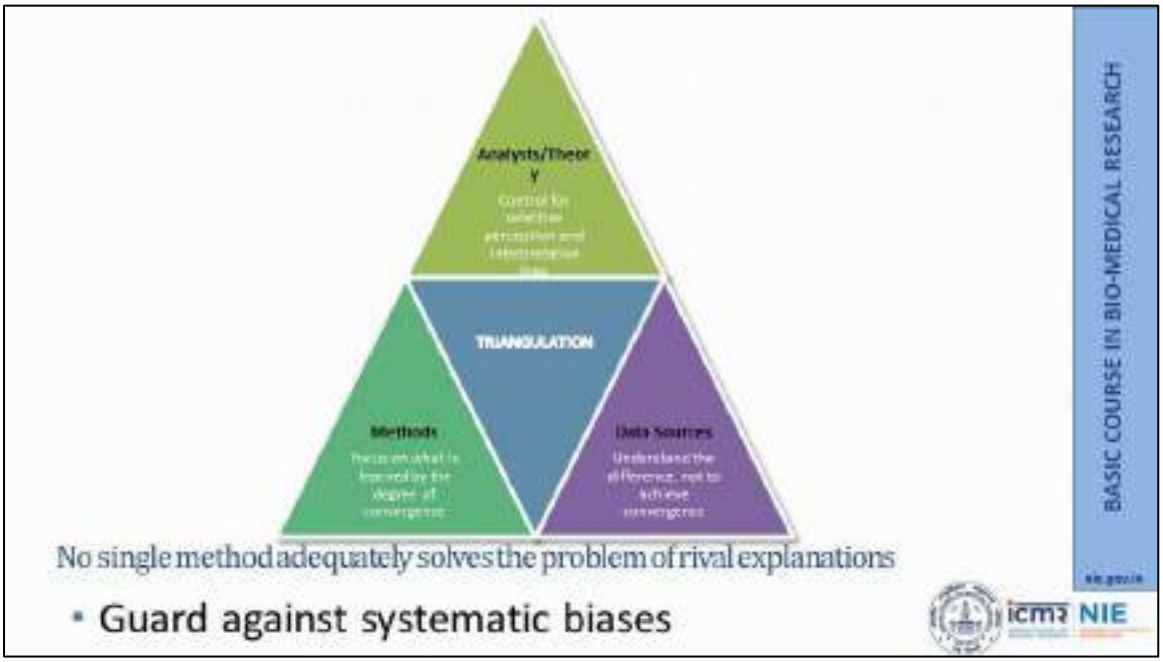
A tool to generate ideas
A preliminary step in developing a quantitative study

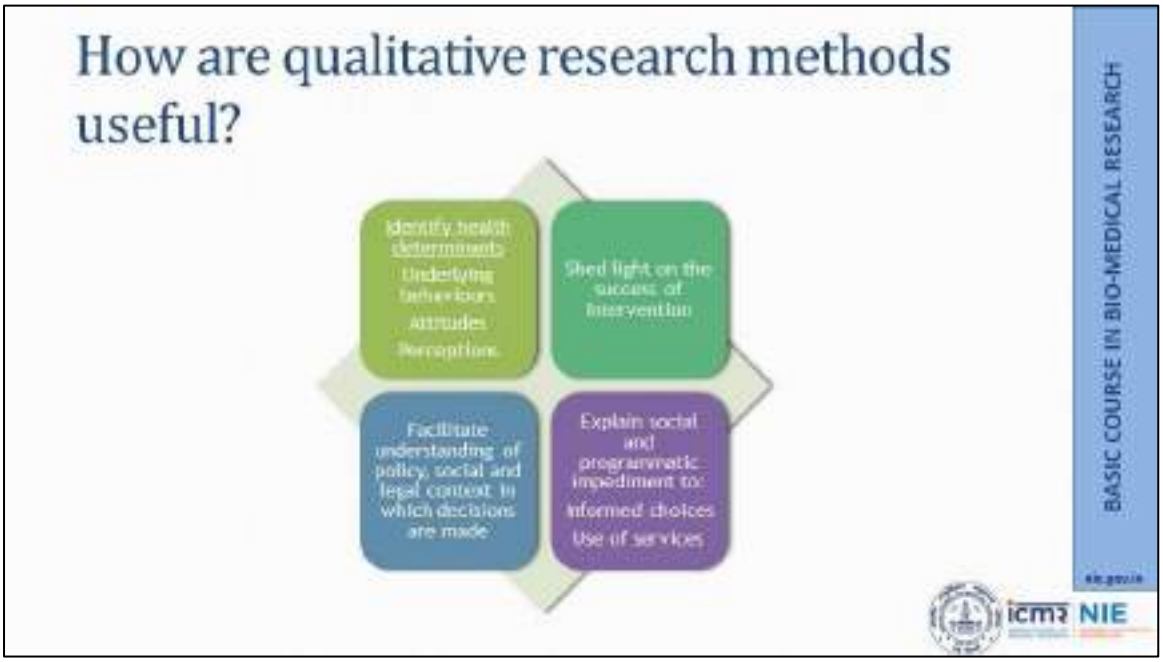
How to use qualitative research methods?

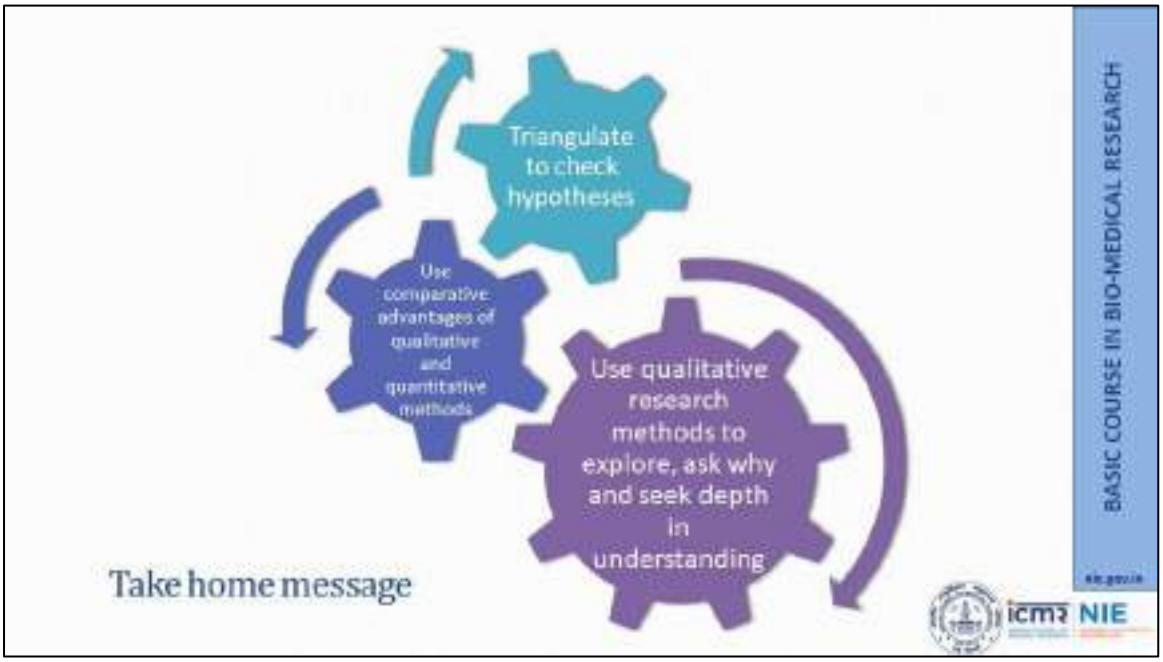
To help understand the results of a quantitative study

How to use qualitative research methods?

The primary data collection method
– sometimes but not necessarily along with a quantitative study like a survey


Analysts/Theor
y
TRIANGULATION
Methods
Data Sources
No single method adequately solves the problem of rival explanations
Guard against systematic biases
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE


How are qualitative research methods useful?
Identify health determinants
Underlying behaviours
Attitudes
Perceptions
Shed light on the success of intervention
Facilitate understanding of policy, social and legal context in which decisions are made
Explain social and programmatic impediment to:
Informed choices
Use of services
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE


Triangulate to check hypotheses
Use comparative advantages of qualitative and quantitative methods
Use qualitative research methods to explore, ask why and seek depth in understanding
Take home message
BASIC COURSE IN BIO-MEDICAL RESEARCH
icmr NIE

1. Which methods in qualitative research use flexible interview guide?
   a) In depth Interviews and participant observation
   b) Focus Group Discussions and In-depth interviews
   c) Participant Observation and focus group discussions
   d) Structure interviews and surveys

2. Which of the following study designs can be used as a tool a generate ideas/hypotheses?
   a) Qualitative study
   b) Case-control study
   c) Experimental study
   d) Cohort study

3. The qualitative data analysis method in which investigators code text into categories and build theoretical models
   a) Content analysis
   b) Grounded theory
   c) Schema analysis
   d) Hermeneutics

4. Open-ended, one-to-one interviews to discover interviewee's own framework of meanings
   a) In-depth Interviews
   b) Focus Group Discussions
   c) Participant observation
   d) Structured interviews

5. Audio recordings during Focus Group Discussions
   a) Can be done without any prior informed written consent
   b) Cannot be done
   c) Should always be done
   d) Can be done with prior informed consent

6. All of the following are situations in which qualitative research methods can be used, EXCEPT
   a) Familiar and sufficiently researched matter
   b) To seek the depth of understanding
   c) Exploration of behaviors
   d) View the social phenomenon holistically

7. The main methods used in qualitative research method include all EXCEPT
   a) In-depth Interviews
   b) Focus Group Discussions
   c) Participant observation
   d) Structured questionnaire-based interviews

8. The observer becomes a part of the group or event in this method of qualitative study
   a) In-depth Interviews
   b) Focus Group Discussions
   c) Participant observation
   d) Structured interviews

9. All the statements regarding Participant Observation is true EXCEPT
   a) Observer becomes a part of the event/group
   b) Systematic collection of data is easy
   c) Analytic methods for observation are not well described
   d) Data is very detailed

10. The qualitative data analysis method which uses theoretical framework as the basis for analysis
    a) Content analysis
    b) Grounded theory
    c) Schema Analysis
    d) Hermeneutics

11. Which of the following are characteristics of qualitative research methods?
    a) Objective, measurable, reliable and repeatable
    b) Subjective, measurable, credible and repeatable
    c) Subjective, credible, inductive and interpretation of responses
    d) Objective, credible, inductive and interpretation of responses

12. The research method which is best suited for collection of information regarding highly sensitive matters such as alcohol use
    a) Focus Group Discussions
    b) Participant Observation
    c) In-Depth Interview
    d) Group discussions

13. Which of the following is not the utility of qualitative research?
    a) To provide insight to why people behave in a certain way
    b) To estimate the prevalence of disease
    c) To help understand the results of a quantitative study
    d) For developing a questionnaire

14. Which of the following statement is true regarding Participant Observation
    a) Observer becomes a part of the event/group
    b) Systematic collection of data is easy
    c) Analytic methods for observation are well described
    d) Data is brief as compared to in-depth interviews

15. Open ended group interviews that promotes discussion among participants is called as
    a) In depth Interviews
    b) Focus Group Discussions
    c) Participant Observation
    d) Structured interviews

16. A researcher decided to conduct a study to explore the child feeding practices among mothers of under five children in a community. Which of the following qualitative techniques can the researcher employ to gather wide range of information on the topic in a short

span of time?
a) Structured interview
b) In depth Interview
c) Participant Observation
d) Focus Group Discussion

17. In qualitative research, researchers interpret the social reality from the participants' point of view.
a) True
b) False

18. Which of the following statements is "Incorrect" about in-depth interviews?
a) Findings are always generalizable
b) The transcripts are time consuming to analyze
c) Helps understand sensitive issues
d) Useful when participants are knowledgeable on a particular topic

19. Which of the following statement is "False" about focus group discussion.
a) Focus group discussions help understand local terminologies
b) Group interaction is integral for an effective discussion
c) Heterogeneity of the group is a pre-requisite
d) Audio and video recordings are done with prior consent

20. Triangulation is the use of multiple methods, multiple theories and or multiple sources for a comprehensive understanding of the topic in question
a) True
b) False

Types of Data

- Qualitative
  - Nominal
    - Eg. Color of Eyes
  - Ordinal
    - Eg. Stages of disease condition
- Quantitative
  - Discrete
    - Eg. Family size
  - Continuous
    - Eg. Height / Weight

Describe – Central Value

- Data is not information.
- Summarize
  - Average
    - Mean
    - Median
    - Mode

Arithmetic Mean (AM)

- Most commonly used; Simply called MEAN
- Add all the observed values (Sum = $\sum X_i$)
- Mean = Sum / n
- Sample Mean is denoted by $\bar{x}$
- Population Mean is denoted by μ

Example

- Age of 10 Pregnant women

26, 31, 25, 21, 26, 26, 27, 25, 27, and 26

Sum = (26+31+25+21+26+26+27+25+27+26) = 260

n = 10

Mean = sum / n = 260/10 = 26 years

Describe – Central Value

- Data is not information.
- Summarize
  - Average
    - Mean
    - Median
    - Mode

Median

The Median describes literally the middle value of the distribution

Divides the distribution exactly into two halves

(i.e. 50% of the data will fall on either side)

Useful when there are extreme values

Example

Duration (days) of hospital stay of 11 patients

1, 2, 3, 4, 5, 6, 7, 8, 8, 9, 77 (Arranged in ascending order)

Median is the middle value (6^{th} value) = 6

(Mean = 11.8)

If n is even; then take average of middle two values.

Describe – Central Value

- Data is not information.
- Summarize
  - Average
    - Mean
    - Median
    - Mode

Mode

The Mode is the value that occurs most frequently

Mode is the only location statistics to be used – for nominal data - not measurable characteristic

Epidemiology – Describing an epidemic with respect to TIME

Example

- Colour preference of people for their car

Colour preference	No. of persons
Green	354
Yellow	852
White	320
Red	474

Mode = Yellow

Describe – Dispersion

- Is it enough to know the average?
  - Example of swimming pool.
- Measures of variability
  - Range
  - Inter-quartile range
  - Mean deviation from mean
  - Variance / Standard deviation

RANGE

Definition:

The difference between the Minimum and the Maximum value of the observations

Advantage:

A quick and easy indicator of dispersion.

Disadvantage:

Influenced by extreme values; Uses only two data points

INTER-QUARTILE RANGE
Quartile Deviation

Definition:

Defined as the interval between the value of the upper quartile (Q3) and the lower quartile (Q1)
Inter Quartile Range = $Q_3 - Q_1$

Advantage:

Unaffected by the extreme values

Disadvantage:

Covers only the middle 50% observations

Describe – Dispersion

- Is it enough to know the average?
  - Example of swimming pool.
- Measures of variability
  - Range
  - Inter-quartile range
  - Mean deviation from mean
  - Variance / Standard deviation

MEAN DEVIATION

Definition: **The mean deviation is the average of the absolute (ignoring the sign) deviations of the observations from the arithmetic mean.**

Advantage: **It is based on all the observations in the group. It is easy to grasp the meaning of the procedure.**

Disadvantage: **It ignores the sign of the difference of the value of the observation and arithmetic mean.**

It is not widely used because of the availability of a more advantageous measure.

Describe – Dispersion

- Is it enough to know the average?
  - Example of swimming pool.
- Measures of variability
  - Range
  - Inter-quartile range
  - Mean deviation from mean
  - Variance / Standard deviation

STANDARD DEVIATION - SD (σ)

Definition: **The SD is the square root of the average of the squared deviations of the observations from the arithmetic mean**

The square of the SD is called variance

Advantage: **The SD is the most important measure of distribution. While the variance is in unit squared, the SD is expressed in the same units of measurement as the observation. It is suitable for further analysis**

The SD together with arithmetic mean is useful for description of the data

Coefficient of Variation (CV)

Purpose: To compare the relative variability in different groups

Definition: The coefficient of variation is the SD expressed as a percentage of the arithmetic mean (AM).

$$CV = \left(\frac{SD}{AM}\right) \times 100$$

Summary

- Choose appropriate central / dispersion value
  - Mean / SD – if no extreme values
  - Median / IQR – if there are extreme values
  - Mode /Range – for qualitative variables/ time distribution in epidemic curve
- Mean and SD are used the most.

1. In a study on hypertension, patients are categorized based on their systolic blood pressure as normal, pre-hypertension, stage 1 hypertension and stage 2 hypertension. What type of variable is this?
   a) Qualitative
   b) Descriptive
   c) Nominal
   d) Ordinal

2. Most commonly used measure of central tendency is
   a) Mode
   b) Median
   c) Mean
   d) Range

3. First quartile (Q1) is equivalent to _____ percentile
   a) 25th
   b) 50th
   c) 75th
   d) 1st

4. Find the median in the following sample of observations: 12, 26, 10, 29, 48
   a) 29
   b) 48
   c) 26
   d) 25

5. The following measure is not influenced by extreme values in a data set
   a) Arithmetic Mean
   b) Inter-quartile range
   c) Range
   d) 'b' and 'c'

6. Which of the following statistic does not belong with the others?
   a) Range
   b) Variance
   c) Mode
   d) Standard deviation

7. 'Number of children per household' is an example of a continuous variable
   a) True
   b) False

8. In a study, researchers are interested in measuring the cholesterol levels of participants. Cholesterol level is a _____ variable
   a) Ordinal
   b) Nominal
   c) Continuous
   d) Discrete

9. In the following set of data, what is the mean? 4,1,9,7,3,8,2,6
   a) 5
   b) 4.5
   c) 9
   d) 8

10. Difference between the minimum value and the maximum value of the observations
    a) Variance
    b) Inter-quartile range
    c) Range
    d) Standard Deviation

11. All the following are measures of dispersion EXCEPT
    a) Mean
    b) Variance
    c) Standard deviation
    d) Range

12. Which percentile is equivalent to the median?
    a) 25th
    b) 50th
    c) 75th
    d) 100th

13. All the following are true for standard deviation (SD) EXCEPT
    a) It is the square root of the average of the squared deviations of the observations from the arithmetic mean
    b) It is the most important measure of dispersion
    c) It is expressed in the same units of measurement as the observation
    d) The square of the standard deviation is called mean deviation

14. A researcher measures fasting blood level of glucose of 100 participants. The mean blood sugar level was measured as 110 mg/dl. The standard deviation was 11 mg/dl. Calculate the coefficient of variance.
    a) 20%
    b) 14%
    c) 10%
    d) 25%

15. A researcher measures the height of 100 school going children for his study. What type of variable is 'height'?
    a) Nominal
    b) Ordinal
    c) Continuous
    d) Discrete

16. A chest physician observed the distribution of forced expiratory volume (FEV1) in 100 Chronic Obstructive Pulmonary disease (COPD) patients and calculated a median value of 2.0 litres. The value of first and third quartile of the distribution was 1.5 litres and 3.0 litres respectively. Based on this data how many patients in the sample are expected to have a FEV1 between 1.5 and 3.0 litres?
    a) 100
    b) 50

c) 25
d) 75

17. The average of the absolute deviations of the observations from the arithmetic mean is known as
a) Variance
b) Inter-quartile range
c) Mean deviation
d) Standard deviation

18. In a study, a researcher was interested in measuring the haemoglobin levels of 10 participants. The values are 10.0, 8.5, 12.0, 14.0, 11.5, 13.5, 9.0, 12.0, 11.3, 7.5. What is the mode of this distribution?
a) 7.5
b) 12.0
c) 10.9
d) 14.0

19. All the following are examples of a nominal variable EXCEPT
a) Gender
b) Age
c) Place of residence
d) Colour of eyes

20. Which of the following is true about inter-quartile range?
a) It describes the middle value of the distribution
b) It divides the distribution into two halves
c) It covers the middle 50% of observations
d) It is affected by the extreme values in the distribution

Definition of sampling

Procedure by which some members of the population are selected as representatives of the entire population

Study population

The study population is the population to which the results of the study will be inferred

The study population depends upon the research question

- How many injections do people received each year in India?
  - Study population: Population of India
- How many needle-sticks health care workers experience each year in India?
  - Study population: Health care workers of India
- How many hospitals have a needle-sticks prevention policy in India
  - Study population: Hospitals of India

The sample needs to be representative of the population in terms of time

- Seasonality
- Day of the week
- Time of the day

The sample needs to be representative of the population in terms of place

- Urban
- Rural

The sample needs to be representative of the population in terms of persons

- Age
- Sex
- Other demographic characteristics

Definition of sampling terms

- Sampling unit (Basic sampling unit, BSU)
  - Elementary unit that will be sampled
    - People
    - Health care workers
    - Hospitals
- Sampling frame
  - List of all sampling units in the population
- Sampling scheme
  - Method used to select sampling units from the sampling frame

Why do we sample populations?

- Obtain information from large populations
- Ensure the efficiency of a study
- Obtain more accurate information

Population

Sample

Practical example

- The Ministry of Health of a country X wants to estimate the proportion of children in elementary schools who have been immunized against childhood infectious diseases
- The task must be completed in one month
- The objective is to estimate the proportion of immunized children

Type of samples

- Non-probability samples
  - Probability of being selected is unknown
  - Convenience samples
    - Biased
    - Best or worst scenario
  - Subjective samples
    - Based on knowledge
    - Time/resources constraints
- Probability samples

Type of samples

- Non-probability samples
- Probability samples
  - Every unit in the population has a known probability of being selected
  - Only sampling method that allows to draw valid conclusions about population

Random sampling in probability samples

- Removes the possibility of bias in selection of subjects
- Ensures that each subject has a known probability of being chosen
- Allows application of statistical theory

Sampling error

- No sample is a perfect mirror image of the population
- Magnitude of error can be measured in probability samples
- Expressed by standard error of mean, proportion, differences…
- Function of:
  - Sample size
  - Variability in measurement

Methods used in probability samples

1. Simple random sampling
2. Systematic sampling
3. Stratified sampling
4. Cluster sampling
5. Multistage sampling

1. Simple random sampling

- Principle
  - Equal chance for each sampling unit
- Procedure
  - Number all units
  - Randomly draw units
- Advantages
  - Simple
  - Sampling error easily measured
- Disadvantages
  - Need complete list of units
  - Does not always achieve best representation

Example of simple random sampling

Numbers are selected at random

2. Systematic sampling

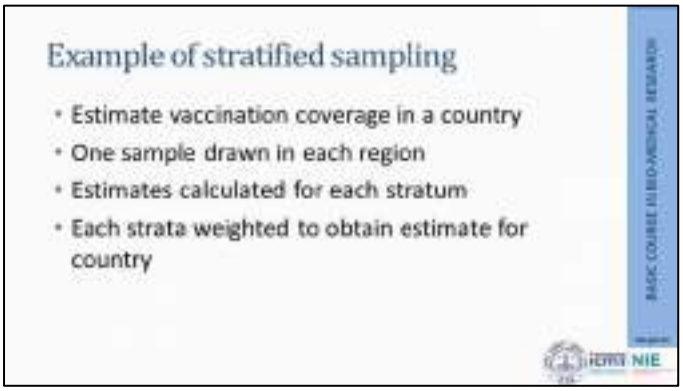
- Principle
  - A unit drawn every k units
  - Equal chance of being drawn for each unit
- Procedure
  - Calculate sampling interval (k = N/n)
  - Draw a random number (≤ k) for starting
  - Draw every k units from first unit
- Advantages
  - Ensures representativity across list
  - Easy to implement
- Disadvantage
  - Dangerous if list has cycles

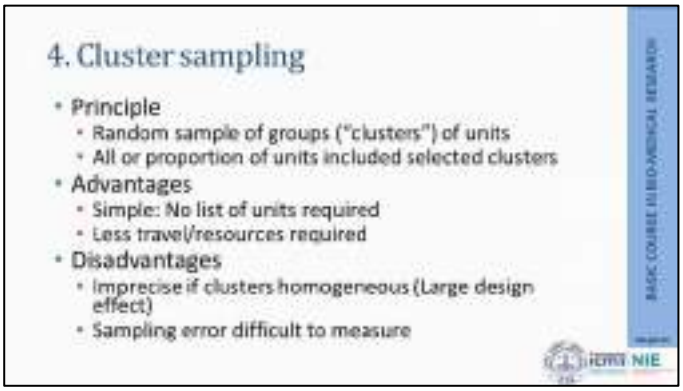
Example of systematic sampling

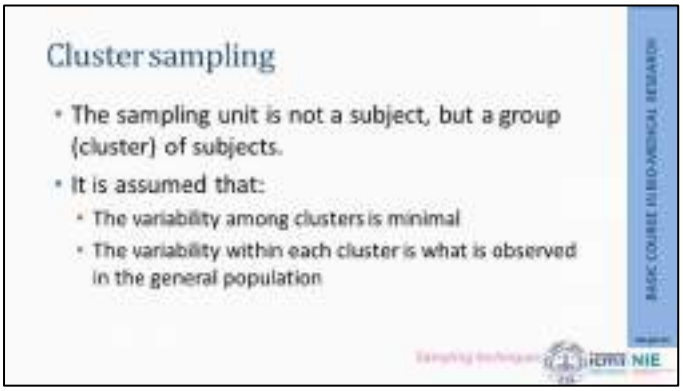
Every eighth house is selected

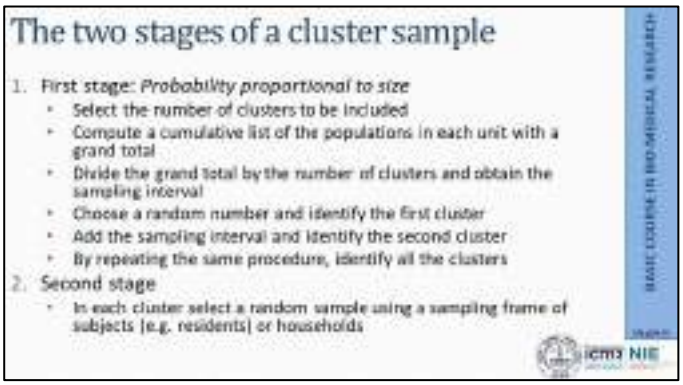
3. Stratified sampling

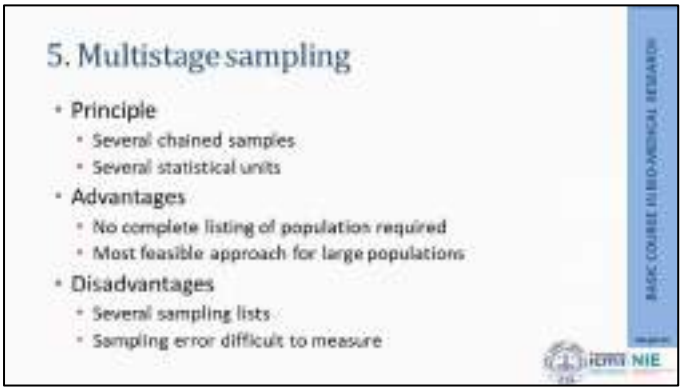
- Principle
  - Classify population into homogeneous subgroups (strata)
  - Draw sample in each strata
  - Combine results of all strata
- Advantage
  - More precise if variable associated with strata
  - All subgroups represented, allowing separate conclusions about each of them
- Disadvantages
  - Sampling error difficult to measure
  - Loss of precision if small numbers sampled in individual strata

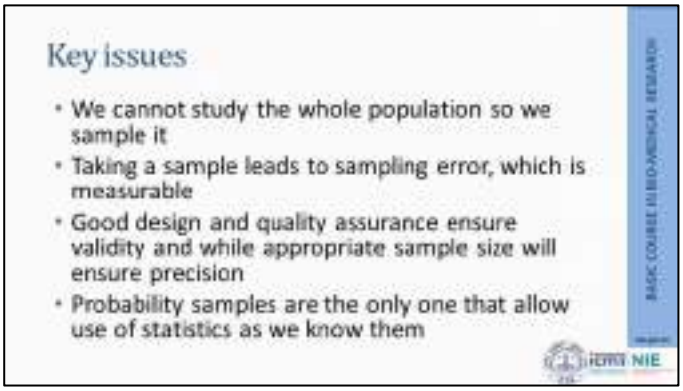












1. The process by which some members of a population are selected as representative of the entire population is known as
   a) Census
   b) Sampling
   c) Survey
   d) Randomization

2. Sampling based upon equal chance of selection is called
   a) Stratified random sampling
   b) Simple random sampling
   c) Systematic sampling
   d) Subjective sampling

3. A researcher wishing to draw a sample from sequentially numbered houses uses a random starting point and then selects every 6th houses, s/he has thus drawn a _____ sample
   a) Sequential
   b) Systematic
   c) Simple random
   d) Stratified

4. The following statement is correct regarding sampling error
   a) Sampling error is difficult to measure in simple random sampling
   b) Sampling error is easy to measure in stratified sampling
   c) The magnitude of error can be measured in non-probability samples
   d) The magnitude of error can be measured in probability samples

5. The only sampling method allows to draw valid conclusions about the population is
   a) Non-probability sampling
   b) Convenience sampling
   c) Probability sampling
   d) Subjective sampling

6. All the following are true regarding cluster sampling EXCEPT
   a) It needs a complete list of units
   b) The sampling unit is group of subjects
   c) Sampling error is difficult to measure
   d) Resources required are less

7. Methods used in probability samples are
   a) Stratified sampling
   b) Multi-stage sampling
   c) Cluster sampling
   d) All of the above

8. All the following statements are true regarding simple random sampling EXCEPT
   a) Sampling error is easily measurable
   b) It needs a complete list of all units
   c) It ensures equal chance of selection for each unit
   d) It always achieves best representativeness

9. People who volunteer or who can be easily recruited are used in a sampling method called
   a) Cluster sampling
   b) Multi-stage sampling
   c) Convenience sampling
   d) Systematic sampling

10. Based on the number of cigarettes per day, a researcher divides the population into three risk groups for lung cancer (low, moderate, high risk). If the researcher then draws a random sample from each of these risk groups independently, s/he has created a _____ sample
    a) Systematic
    b) Simple random
    c) Stratified
    d) Group data

11. All the following are non-probability sampling methods EXCEPT
    a) Convenience sampling
    b) Snowball sampling
    c) Quota sampling
    d) Systematic sampling

12. In a study to measure the prevalence of fluorosis in a district, towns are sampled first. This is followed by a sample of wards within the selected towns, and finally a sample of households within the selected wards. What is the type of the sampling used here?
    a) Multistage sampling
    b) Systematic random sampling
    c) Simple random sampling
    d) Convenience sampling

13. The magnitude of sampling error can be measured in probability sampling.
    a) True
    b) False

14. All the following statements are true regarding stratified sampling EXCEPT
    a) It classifies population into homogeneous subgroups
    b) The probability of a participant being selected is unknown
    c) The sampling error is difficult to measure
    d) It allows inclusion of representative participants from all subgroups

15. Which of the following is true about non-probability sampling?
    a) It removes the possibility of bias in selection of participants
    b) Sampling error can be measured
    c) Quota sampling is a type of non-probability sampling
    d) Inferences drawn from non-probability sampling can be generalized

16. Random sampling in probability samples reduces the possibility of selection bias
    a) True
    b) False

17. Which of the following statement is true regarding systematic random sampling?
    a) Sampling error cannot be measured
    b) The chance of selection for each sampling unit is unknown
    c) The selected sampling units are likely to be more representative than simple random sampling
    d) It is a type of non-probability sampling

18. A researcher planned a cross-sectional study to assess the level of satisfaction of patients attending a clinic. For this, the researcher selected the first 100 patients who visited the clinic starting from a fixed date. What is the type of the sampling mentioned in this case?
    a) Snowball sampling
    b) Purposive sampling
    c) Simple random sampling
    d) Stratified random sampling

19. The list of all individuals in the study population from whom study participants in a research are to be selected is known as
    a) Sampling frame
    b) Study population
    c) Sampling unit
    d) Study sample

20. Which of the following is an advantage of multistage sampling?
    a) Sampling error is easy to measure
    b) It does not require a complete list of the total population
    c) It requires only one sampling list
    d) It always achieves the best representative sample

Objectives

- Understand the relationship between sample size and power
- Determine sample size necessary to achieve a given level of power for estimating a simple proportion, and other measures of effect

Steps in Estimating Sample Size

- Identify major study variable
- Determine type of estimate (%, mean, ratio,...)
- Indicate expected frequency of factor of interest
- Decide on desired precision of the estimate
- Decide on acceptable risk that estimate will fall outside its real population value
- Adjust for population size
- Adjust for estimated design effect
- Adjust for expected response rate

α and Confidence Level

- α: The significance level of a test: the probability of rejecting the null hypothesis when it is true (or the probability of making a Type I error).
- Confidence level: The probability that an estimate of a population parameter is within certain specified limits of the true value; commonly denoted by "1- α".

β and Power

- β: The probability of failing to reject the null hypothesis when it is false (or the probability of making a Type II error).
- Power: The probability of correctly rejecting the null hypothesis when it is false; commonly denoted by "1- β"

Precision

A measure of how close an estimate is to the true value of a population parameter. It may be expressed in absolute terms or relative to the estimate.

Sample Size Required for Estimating Population Mean

- Suppose we want an interval that extends d units on either side of the estimator
  d = (reliability coefficient) x (Standard error)
- If sampling is from a population sufficiently large size, the equation is:
  $$d = z\frac{\sigma}{\sqrt{n}}$$
- When solved for n gives:
  $$n = \frac{z^2 \cdot \sigma^2}{d^2}$$

Example 1 (1/2) What Sample Size Do I Need If...?

A health department nutritionist , wishing to conduct a survey among a population of teenage girls to determine the average daily protein intake

What information is needed to estimate the sample size?

- The nutritionist must provide three items of information: the desired width of the confidence interval, the level of confidence desired, and the magnitude of the population variance

Example 1 (2/2) What Sample Size Do I Need If...?

- Solution: The nutritionist would like an interval about 10 units wide; that is, the estimate should be within about 5 units of the true value in either direction. A confidence coefficient of .95 is decided and on that, from past experience, the nutritionist feels that the population standard deviation is probably about 20 grams.
- Summarizing the information: $z = 1.96$, $\sigma = 20$, and $d = 5$
- Calculation:
  $$n = \frac{(1.96)^2 (20)^2}{(5)^2} = 61.47$$

A note on Population Standard Deviation σ

- The formulas for sample size require knowledge of σ^2. However, in general, the population variance is unknown and has to be estimated:
  - A pilot or preliminary sample. Observations used in the pilot can be counted as part of the final sample
  - Estimates may be available from previous studies
  - If thought that the population is approximately normally distributed, we may use the fact that the range (R) is approximately equal to 6 standard deviations.

  $\sigma \approx R/6$

Sample Size Required for Estimating Proportions

- The formula requires the knowledge of p, the proportion in the population possessing the characteristic of interest. However, this is what we are trying to estimate and is unknown
  - A pilot or preliminary sample. Observations used in the pilot study can be counted as part of the final sample
  - Estimates may be available from previous studies and the upper bound of p can be used in the formula
  - If impossible to come with a better estimate, set $p = 0.5$ in the formula to yield the maximum value of n

Sample Size Required for Estimating Proportions

The method is essentially the same as for population mean. Assuming random sampling and approximate normality in the distribution of p, brings us to the formula for n if sampling is with replacement, from a population sufficiently large to warrant ignoring the finite population correction :

$$n = \frac{z^2 pq}{d^2}$$

Where $q = 1 - p$

Example 2 (1/3)
What Sample Size Do I Need If . . . ?

- I want to estimate the true immunization coverage in a community of school children
- Previous studies tell us that immunization coverage should be somewhere around 80%
- Precision (absolute): we'd like the result to be within 4% of the true value
- Confidence level: conventional = 95% = 1 - α; therefore, $\alpha = 0.05$ and $z_{(1-\alpha/2)} = 1.96$ = value of the standard normal distribution corresponding to a significance level of 0.05 (1.96 for a 2-sided test at the 0.05 level)

Example 2 (2/3)

- d = absolute precision = 0.04
- p = expected proportion in the population = 0.80
- $z_{(1-\alpha/2)} = 1.96$ = value of the standard normal distribution corresponding to a significance level of α (1.96 for a 2-sided test at the 0.05 level)

Example 2 (3/3)
Sample Size

$$n = \frac{z^2 \cdot p \cdot (1-p)}{d^2}$$

$$= \frac{(1.96)^2 (.80)(.20)}{(0.04)^2}$$

$$= 384$$

Design Effect

- A bias in the variance introduced in the sampling design, by selecting subjects whose results are not independent from each other; relative change (increase) in the variance due to the use of clusters.
- The design effect can be calculated after study completion, but should be accounted for at the design stage.
  - The design effect is 1 (i.e., no design effect) when taking a simple random sample.
  - The design effect varies using cluster sampling; it is usually estimated that the design effect is 2 in immunization cluster surveys.

What You Need to Calculate Sample Size for Analytical Studies

- Desired values for the probabilities of α and β
- The proportion of the baseline (controls or non-exposed) population
  - EXPOSED (for case-control studies), or
  - DISEASED (for cohort/intervention studies)
  - Often based on previous studies or reports
- Magnitude of the expected effect (RR, OR)
  - Often based on previous studies or reports
  - Minimum effect that investigator considers worth detecting
- Formula: different formulae depending on study design, research question, and type of data

Example 3 (1/3)
What Sample Size Do I Need If . . . ?

- Cohort study of oral contraceptive (OC) use in relation to risk of MI among women of childbearing age
- Previous studies
  - Proportion of non-OC users who are at risk of disease = 0.15
  - Proportion of OC-users who are at risk of disease = 0.25
- Conventional $\alpha = 0.05$ (two-sided)
- Conventional $\beta = 0.20$ (80% power to detect a difference if one truly exists)
- Assume equal sample sizes ($n_1 = n_2$)

Example 3 (2/3)

- p_0 = proportion of non-OC users who are diseased = 0.15
- p_1 = proportion of OC-users who are diseased = 0.25
- $q_0 = (1-p_0) = 1.0 - 0.15 = 0.85$
- $q_1 = (1-p_1) = 1.0 - 0.25 = 0.75$
- $z_{(1-\alpha/2)} = 1.96$ = value of the standard normal distribution corresponding to a significance level of α (1.96 for a 2-sided test at the 0.05 level)
- $z_{(1-\beta)} = 0.84$ = value of the standard normal distribution corresponding to the desired level of power (0.84 for a power of 80%)

Example 3 (3/3)

$$n \text{ (each group)} = \frac{(p_0 q_0 + p_1 q_1)(z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_0)^2}$$

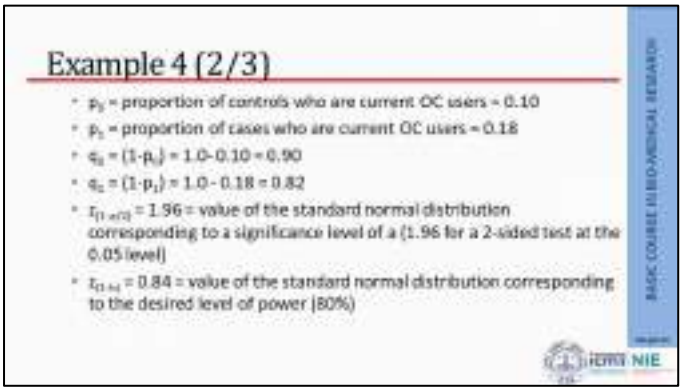
$$\frac{[(.15)(.85) + (.25)(.75)](1.96 + 0.84)^2}{(0.25 - 0.15)^2}$$

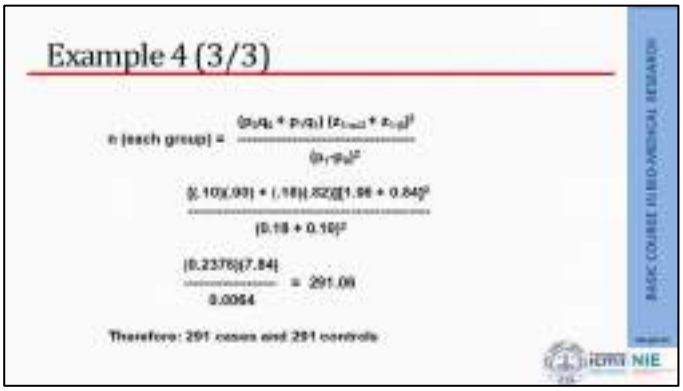
$$\frac{(0.315)(7.84)}{0.01} = 246.96$$

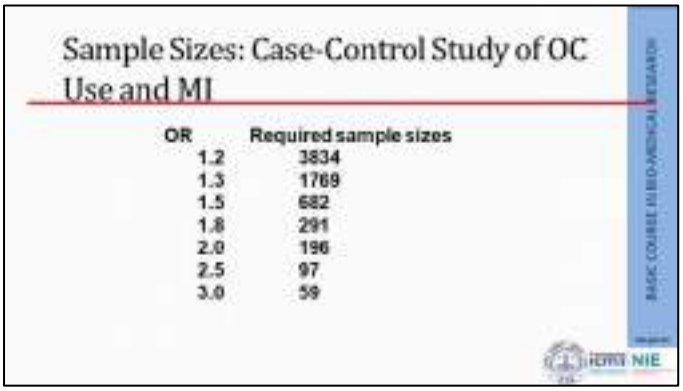
Therefore: 247 OC users (and 247 non-OC users)

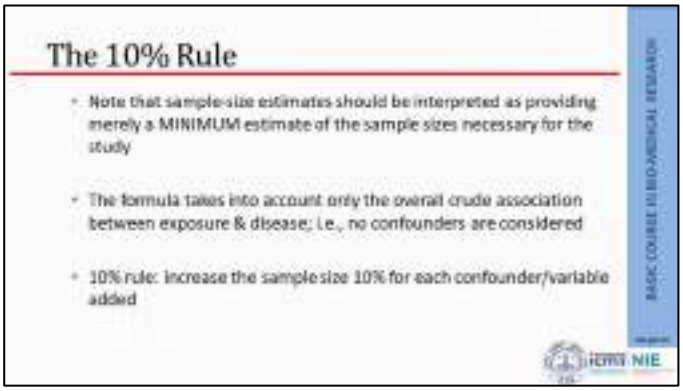
Example 4 (1/3)
What Size Sample Do I Need If . . . ?

- Case-control study of oral contraceptive (OC) use in relation to risk of MI among women of childbearing age
- Previous studies: 10% of women use OCs
- OR of MI associated with current OC use = 1.8
- Conventional $\alpha = 0.05$ (two-sided)
- Conventional $\beta = 0.20$ (80% power to detect difference if one truly exists)
- Assume equal sample sizes ($n_1 = n_2$)









1. Statistical power is defined as the probability of
   a) Accepting a null hypothesis when it is false
   b) Rejecting a null hypothesis when it is true
   c) Rejecting a null hypothesis when it is false
   d) Failing to reject a null hypothesis when it is false

2. Steps in the estimation of sample size included all of the following EXCEPT
   a) Identify major study variable
   b) Decide on the desired precision of the estimate
   c) Adjust for population size
   d) Adjust for selection bias

3. A type-II error occurs when
   a) The null hypothesis is rejected when it is false
   b) The null hypothesis is not rejected when it is false
   c) The null hypothesis is not rejected when it is true
   d) The null hypothesis is rejected when it is true

4. Exact calculation of design effect for a study parameter can take place you after study completion
   a) True
   b) False

5. Population variance can be estimated from
   a) A pilot study
   b) Reports of previous studies
   c) Guessing
   d) 'a' and 'b'

6. The recommended minimum level of power for an analytical study
   a) 5%
   b) 95%
   c) 80%
   d) 0.05%

7. In general, sample size formula takes into account the crude association between exposure and outcome as well as the confounders
   a) True
   b) False

8. Design effect of 'more than 1' needs to be considered in studies involving
   a) Cluster sampling
   b) Simple random sampling
   c) Stratified random sampling
   d) Non-probability sampling

9. Which of the following is necessary in sample size determination?
   a) Desired confidence level
   b) Desired precision
   c) Magnitude of the population variance
   d) All of the above

10. Which one of the following statements is false?
   a) Design effect is a relative change in the variance due to use of clusters
   b) As the magnitude of the expected effect increases, the required sample size increases
   c) The population variance is unknown in general and has to be estimated
   d) Larger the sample size, smaller the sampling error

11. A type-I error occurs when
   a) The null hypothesis is rejected when it is false
   b) The null hypothesis is not rejected when it is false
   c) The null hypothesis is not rejected when it is true
   d) The null hypothesis is rejected when it is true

12. Which of the following is true about β error?
   a) It is the probability of correctly rejecting the null hypothesis when it is false
   b) It is the probability of accepting the null hypothesis when it is false
   c) It is the probability of rejecting the null hypothesis when it is true
   d) It is the probability of making a Type I error

13. All the following are essential statistical considerations for sample size calculation EXCEPT
   a) Desired precision
   b) Anticipated proportion of factor of interest
   c) Sampling method
   d) Allocated budget

14. For each confounder/variable added in the study empirically 10% increase in the sample size should be made.
   a) True
   b) False

15. The design effect should be calculated after completion of the study and it, need not be counted at the design stage.
   a) True
   b) False

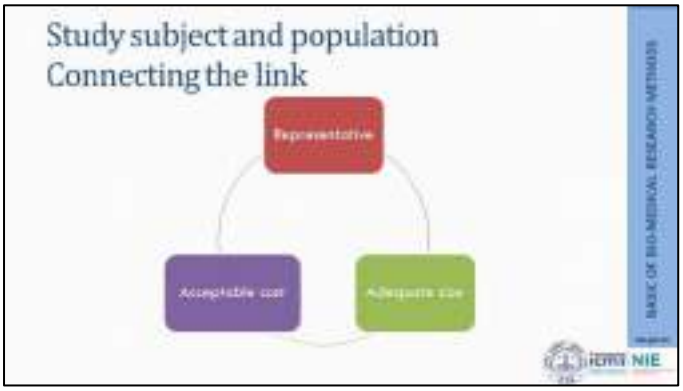
16. When estimating sample size for a cross-sectional study, we need to account for
   a) Expected proportion of characteristic of interest
   b) Estimated design effect, in case of cluster sampling
   c) Population size
   d) All the above

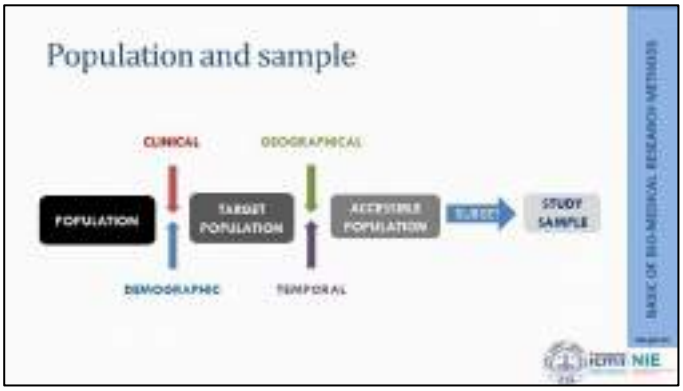
17. The power of a study
   a) Does not influence the sample size
   b) Represented as 'α'
   c) Can be defined as the probability of correctly rejecting null hypothesis when it is false
   d) Represented as the probability of making a Type I error

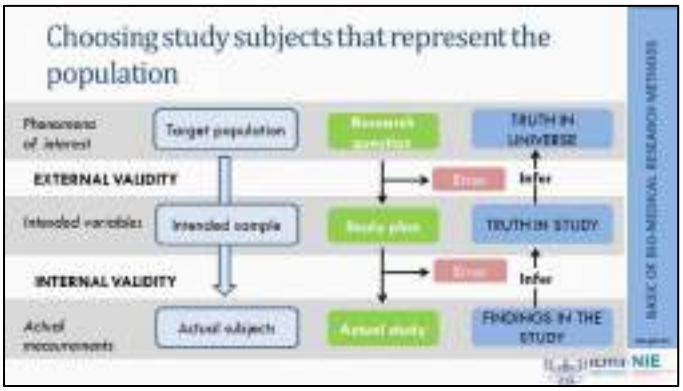
18. The following are needed to calculate sample size for analytical studies using simple random sampling method EXCEPT
   a) Desired value for the probability of α
   b) Magnitude of the expected effect based on previous studies
   c) Desired value for the probability of β
   d) Estimated design effect

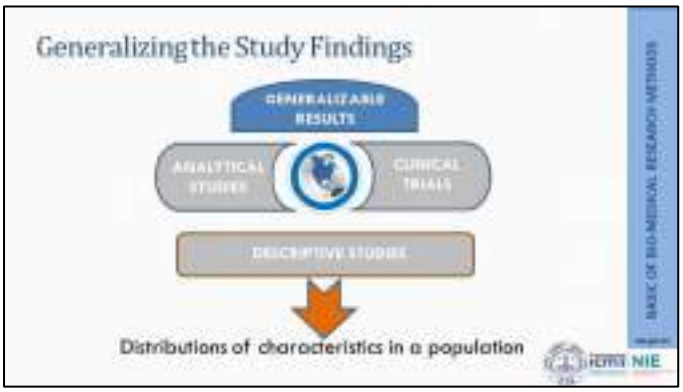
19. A researcher wants to estimate the prevalence of surgical site infection following cesarean section at a tertiary care hospital. What would be the minimum number of sample size to estimate the magnitude of surgical site infection following cesarean section if it is estimated that the proportion of surgical site infection will be 10% in the hospital considering 5% absolute precision and 95% confidence level (Z α/2 = 1.96).
   a) 100
   b) 138
   c) 148
   d) 158

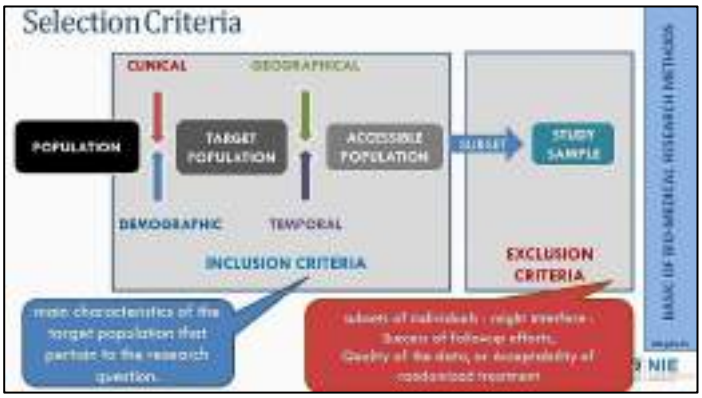
20. Precision is described as a measure of how close an estimate is to the true value of a population parameter.
   a) True
   b) False









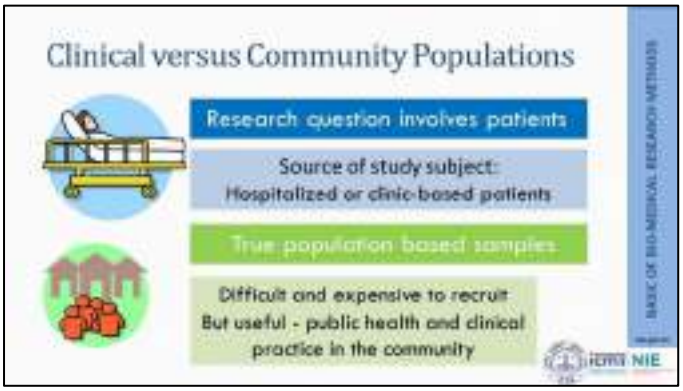


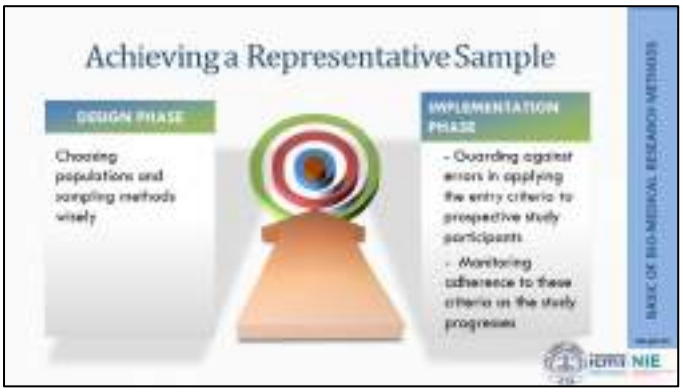
Selection Criteria - example

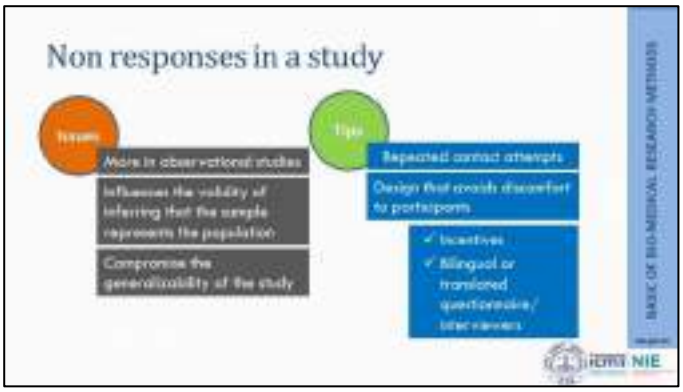
Designing Selection Criteria for a Clinical Trial of Low Dose Metformin to reduce dysmenorrhea in females with poly cystic ovary

Inclusion criteria (Specifying populations relevant to the research question and efficient for study)	Demographic characteristics	Females in reproductive age group (15 -44 years)
	Clinical characteristics	Females with Poly cystic ovary with dysmenorrhea
	Geographic characteristics	Patients attending OPD of the hospital in that region
	Temporal characteristics	Between Jan 1 – Dec 31 of specified year
Exclusion criteria (Subset of population will not be studied because of)	Interfere with loss to follow-up	Chances of moving out of location or marriage
	Interfere with quality of data	Patients already on metformin therapy for other cause
	Being at high risk of possible adverse effects	Hypersensitivity to metformin / Renal dysfunction

BASIC OF BIO-MEDICAL RESEARCH METHODS
NIE







1. Selection of study participants depends on
   a) Representativeness
   b) Acceptable cost
   c) Adequate size
   d) All of the above

2. Target population is determined by
   a) Demographic characteristics
   b) Temporal characteristics
   c) Clinical characteristics
   d) 'a' and 'c'

3. Study sample is a subset of accessible population
   a) True
   b) False

4. Representativeness of a study sample refers to
   a) The extent to which the characteristics of the sample accurately reflect the characteristics of the population
   b) The size of the sample which is large enough
   c) Volunteering nature of the subjects from the population
   d) The extent to which the characteristics of exposed population accurately reflect the characteristics of unexposed sample

5. Non-response in a study can be minimized by
   a) Repeat contact of the study participants
   b) Providing compensation for participants time
   c) Less invasive and less sensitive questionnaires
   d) All of the above

6. External validity means
   a) The degree to which the inferences drawn from a study can be generalized to a broader population beyond the study population
   b) The degree to which the observed findings lead to correct inferences about phenomena taking place in the study sample
   c) The degree to which a test actually measures what it is designed to measure
   d) The degree to which the findings are reliable

7. Participants may be excluded from the study because of
a) Interference with the success of study follow-up
b) Ethical concerns
c) Interference with the quality of data collection or non-acceptance to participate in the study
d) All of the above

8. While choosing the accessible population and the sampling approach for selection of study population, an important factor that we need to consider is
a) Simplicity
b) Technology
c) Feasibility
d) Reliability

9. If your research question is related to diagnosis, treatment or prognosis of a severe medical condition, then it is an easy and cost-effective way to recruit the study population from the community
a) True
b) False

10. Reasons for interference with the success of follow-up in a study may include
a) Migration of some study participants from the study area
b) Marriage of some of the female study participants because of which they might move out of the study area
c) Refusals for follow-up
d) All of the above

11. The population defined by clinical and demographic characteristics is called
a) Target population
b) Accessible population
c) Subset
d) Study sample

12. The population defined by geographical and temporal characteristics is called as
a) Target population
b) Accessible population
c) Subset
d) Sample size

13. Random errors can be effectively handled by
a) Randomisation
b) Representativeness
c) Adequate sample size
d) All of the above

14. A researcher found an inference about a particular disease of interest. If he/she wants to generalize the results, it is important to have
a) Internal validity
b) External validity
c) Feasibility
d) Accuracy

15. Reasons for interference with the success of follow-up in a study may include
a) Out-migration of some study participants from the study area
b) Marriage of some of the female study participants because of which they might move out of the study area
c) Refusals for follow-up
d) All of the above

16. Less invasive and less sensitive questionnaires will
a) Increase the power
b) Decrease the power
c) Improve the significance
d) Reduce the non-response

17. The external validity in a research study means
a) The degree to which the observed findings lead to correct inferences about phenomena taking place in the study sample
b) The degree to which a test actually measures what it is designed to measure
c) The degree to which the inferences drawn from a study can be generalized to a broader population beyond the study population
d) The degree to which the findings are reliable

18. The degree to which the observed findings lead to correct inferences about phenomena of interest in the study sample is
a) Reliability
b) Feasibility
c) Internal validity
d) External validity

19. Which factor is important to consider while choosing the accessible population and the sampling approach?
a) Feasibility
b) Sensitivity
c) Specificity
d) Reliability

20. The participants may be excluded from the study because of
a) Interference with the success of study follow-up
b) Ethical concerns
c) Interference with the quality of data collection
d) All of the above

Principles of project management

To ensure that the defined objectives are met

To also ensure that products/ deliverables are delivered within the defined timeframe and budget at the expected quality standards

The end result should be to provide directions for future implications .. Basically for better tomorrow

BASIC OF BIO-MEDICAL RESEARCH METHODS

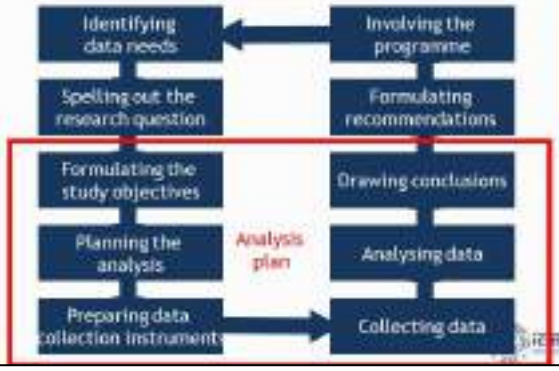
Principles of project management

Ad hoc approach to conducting a research study is often non-productive

The confusion at the beginning of the study
- I want to do a study but I am not clear about the objectives
- I have prepared a questionnaire, but I am not clear about exact information I need
- I will collect data, but I am not clear how I will use that

The disastrous end result.....
- I have data that are difficult to analyse
- I have analysed the data but finding it difficult to interpret
- The interpretations are difficult to use in programs or for policy making

BASIC OF BIO-MEDICAL RESEARCH METHODS



A road map to study planning and management

- Formulate appropriate objectives for the study
- Choose the right design to determine key indicators
- Identify parameters needed for the key indicators
- Prepare the analysis outline
- Estimate sample size

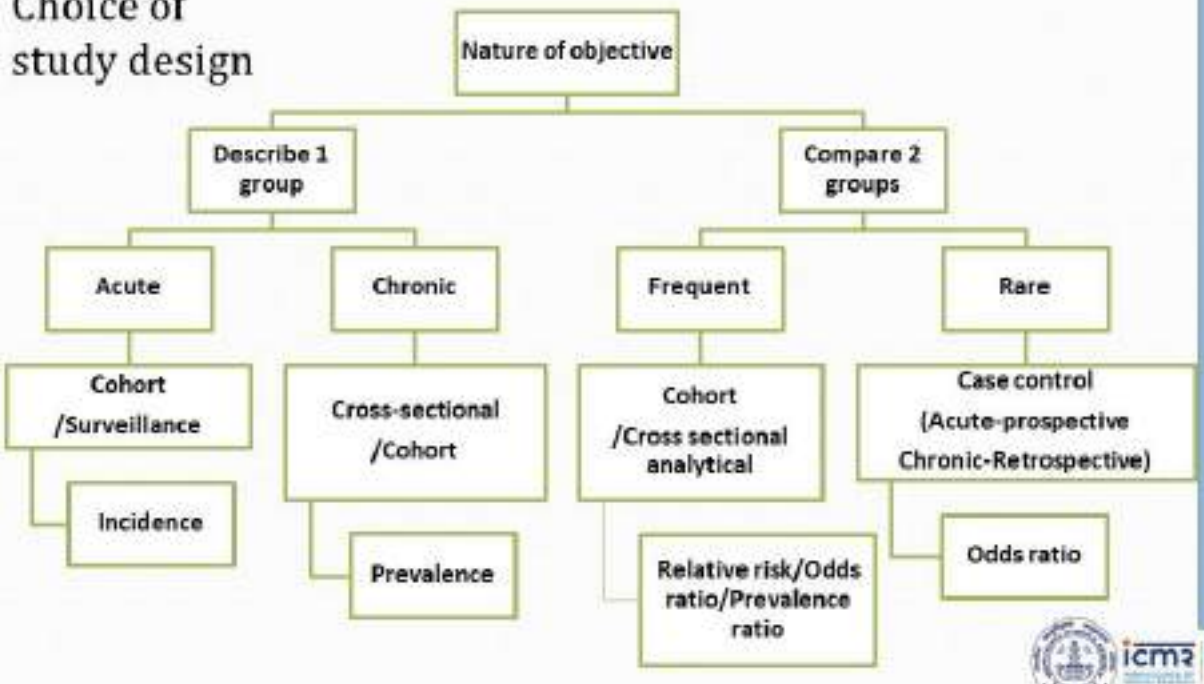
BASIC OF BIO-MEDICAL RESEARCH METHODS

Framing the study objectives is critical

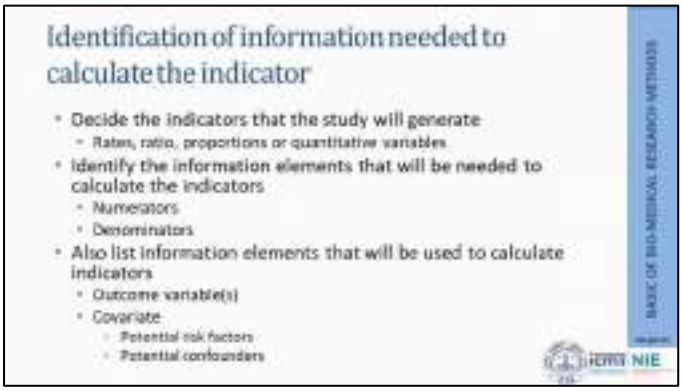
- Fewer the better ..
- May be mentioned as primary and secondary
- Should be clearly phrased:
  - Aimed at testing a hypothesis: **Determine** whether a contaminated well caused an outbreak
  - Aimed at measuring a quantity: **Estimate** the prevalence of diabetes

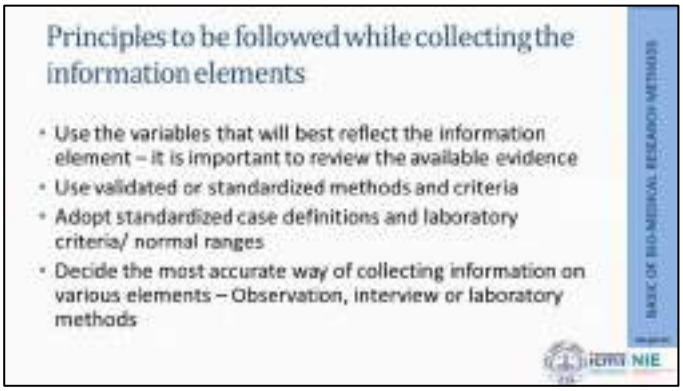
BASIC OF BIO-MEDICAL RESEARCH METHODS

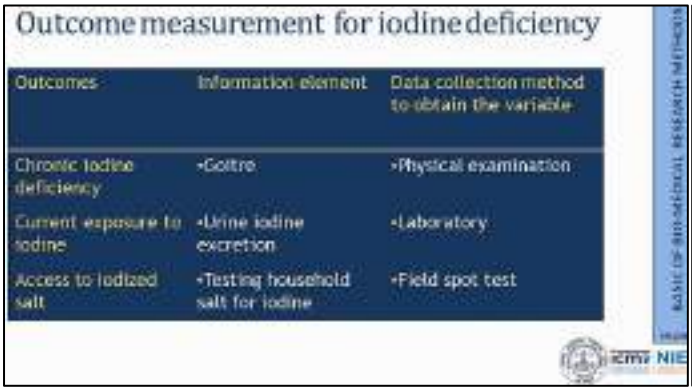
Choice of study design

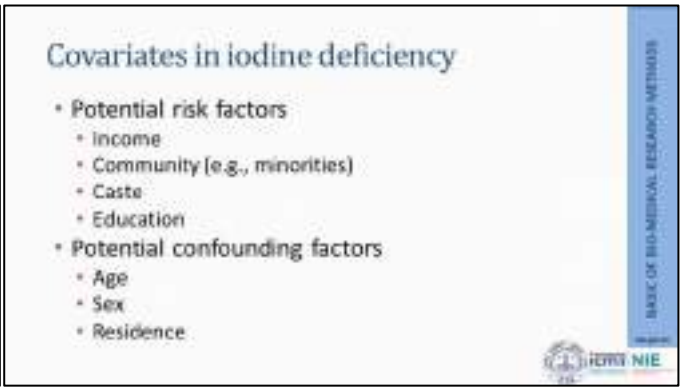


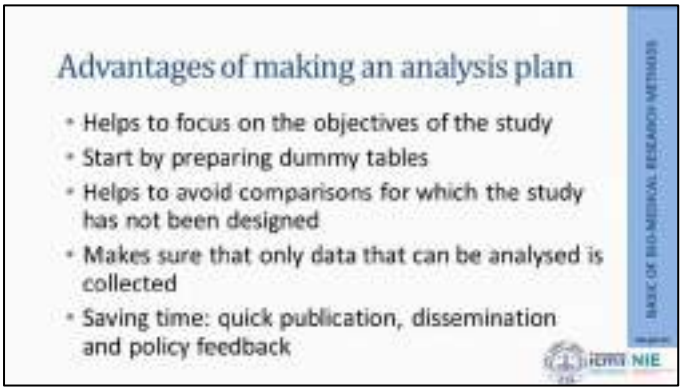
BASIC OF BIO-MEDICAL RESEARCH METHODS

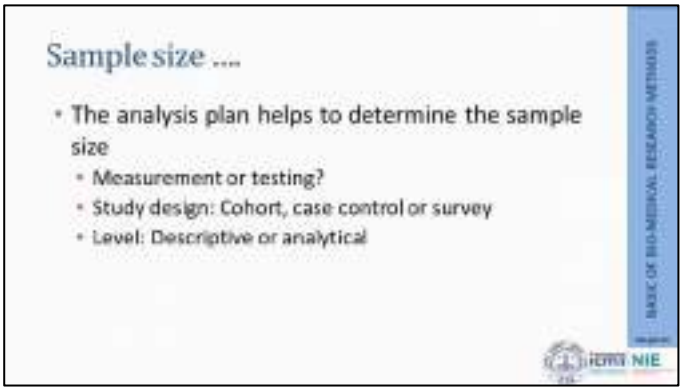


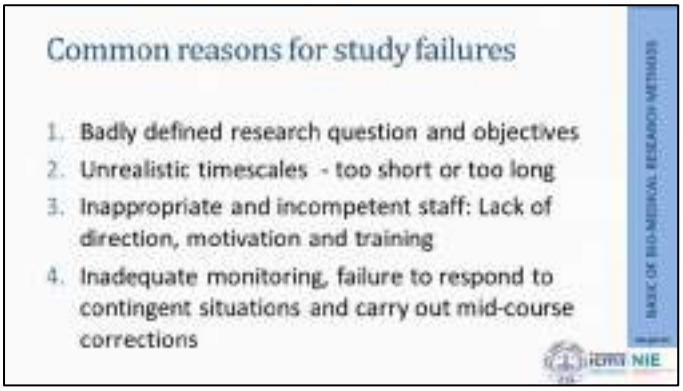












1. Which of the following statements regarding study objectives is correct?
   a) Objectives should be defined at the planning stage of study
   b) Objectives can be defined at any time of the study
   Objectives can be changed even at the end of the study
   d) Objectives should be defined before identifying the research question

2. Which of the following is (are) required to determine the key indicators for planned research study?
   a) Frame study objectives
   b) Identify parameters needed for the key indicators
   c) Choose the right study design
   d) All of the above

3. Which of the following can improve efficiency of a research study?
   a) Time management
   b) Planning and scheduling activities
   c) Budgeting
   d) All of the above

4. Which of the following represents the correct sequence in a life cycle of a study?
   a) Identifying data needs, formulating study objectives,

planning analysis, spelling out research question
b) Formulating study objectives, planning analysis, spelling out research question, identifying data needs
c) Identifying data needs, spelling out research question, formulating study objectives, planning analysis
d) Formulating study objects, spelling out research question, identifying data needs, planning analysis

5. It should be ensured that products/deliverables of health research projects are delivered within the
a) Defined timeframe
b) Defined budget
c) Expected quality standards
d) All of the above

6. Which of the following statements best describes the study objectives?
a) They should be minimum, achievable and clear
b) They can be primary and / or secondary
c) Adding objectives during study implementation is a good practice
d) 'a' and 'b'

7. Principles to be followed while collecting the information elements are
a) Use the variables that will best reflect the information element
b) Adopt standardize case definitions and laboratory criteria/normal ranges
c) Choose the most accurate ways of collecting information on various elements
d) All of the above

8. Study conducted following an *ad hoc* approach may lead to the following consequences
a) Generation of useful data in programs or for policy making
b) Efficient utilization of resources
c) Serious difficulties in analysis and interpretation
d) All of the above

9. Common reasons for research study failures
a) Poorly defined research question
b) Vague timelines
c) Lack of supervision
d) All of the above

10. Sample size for a cross-sectional study is decided based on the following
a) Assumed/reported prevalence
b) Confidence interval
c) Acceptable precision
d) All of the above

11. Which of the following is a criterion for a good research question?
a) Long and self-explanatory question using complex terms
b) A question based on ill-defined hypothesis
c) A question based on strong hunch on part of the investigator
d) A question based on established theory and some research evidence

12. Which of the following can be considered true in case of ad hoc approach to conduct a research study?
a) Its advantages are the low development effort and possibility of getting results in a short time span
b) Its advantage is that the accuracy of the results is usually high
c) Only 'a'
d) Both 'a' and 'b'

13. A cross sectional study is carried out to examine whether naval medical personnel of a higher rank have more positive copying skills than those of a lower rank. Which of the following statement is true of this study?
a) Neither variable is dependent as the researcher cannot manipulate them
b) The independent variable is rank and the dependent variable is copying skills
c) The independent variable is copying skills and the dependent variable is rank
d) None of the above

14. Indicators are considered positive when they have a direct relationship (association, correlation) with the state of health. Which of the following are the examples of positive indicators?
a) The proportion of cured tuberculosis cases
b) Incidence of AIDS
c) Life expectancy at birth
d) 'a' and 'c'
e) 'b' and 'd'
f) 'a', 'b', 'c' and 'd'

15. Which of the following techniques is preferentially used when the population is finite?
a) Purposive sampling technique
b) Area sampling technique
c) Systematic sampling technique
d) None of the above

16. A study began in 1980 with enrollment of a group of 7000 adults in Pondicherry who were asked about their alcohol consumption, smoking, diet, environmental risk factors etc. All the participants were periodically examined and evaluated for evidence of various types of cancers between 1990-1995. Which of the following study designs was used by the investigators?
a) Case-control study
b) Prospective cohort study
c) Ecological study
d) Retrospective cohort study

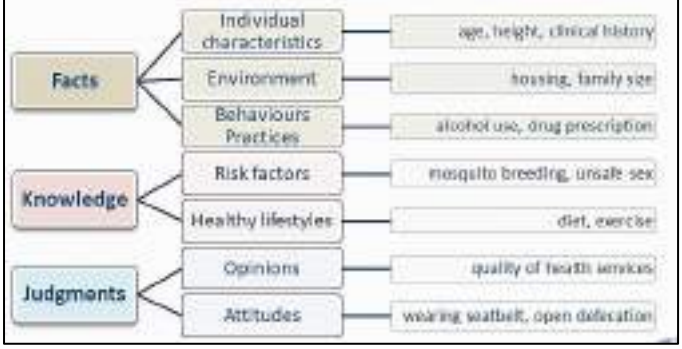
17. An increased number of postoperative wound infections were recorded in patients who underwent incision appendectomy compared with those who had a laparoscopic procedure. Which of the following statement/s is/are true in such a scenario?
    a) This association may simply be owing to the presence of a confounding factor
    b) Association between the two can be better studied in randomized controlled clinical trials
    c) Both 'a' and 'b'
    d) None

18. In a study to evaluate the effectiveness of a new medication, which of the following will generate a stronger evidence?
    a) Comparing outcomes among those receiving medication with those not receiving the same.
    b) Comparing outcomes among those receiving higher doses of medication with those receiving lower doses
    c) Comparing adverse events and drug reactions among those receiving medication and those not receiving medication
    d) All of the above

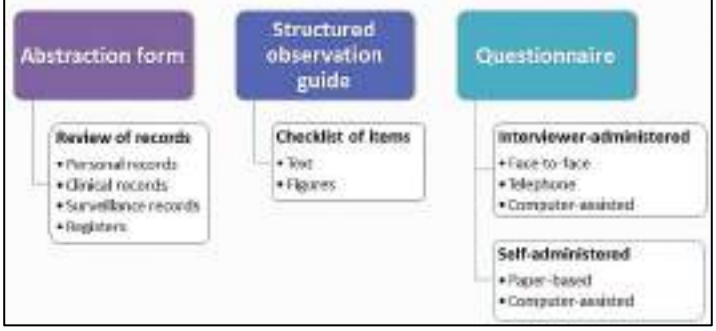
19. Smart objectives are goals that are designed to be be specific, measurable, achievable, relevant and time-bound. Which of the following is an illustration of non-measurable objective?
    a) Incidence of colorectal cancers in Indian adult men
    b) Experiences shared by victims of domestic violence
    c) To determine if regular skin emollients applied from 2 weeks of age reduced development of atopic dermatitis by age 12 months in the general infant population
    d) None of the above

20. Validity of a research can be improved by:
    a) Taking the true representative sample of the population
    b) Eliminating extraneous factors and collecting detailed information on confounding factors
    c) 'a' and 'b'
    d) None of these

Information collected with data collection tools

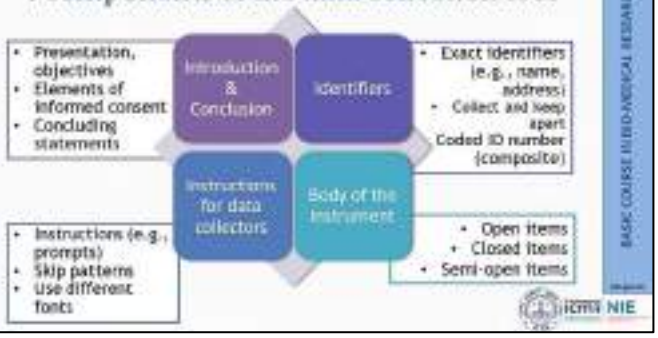


Different tools to collect data



Key elements of data collection tools

4 components of the data collection tool



Open questions

- Answers are not suggested
- Respondents must generate an answer
- Advantages
  - Give freedom of response
  - Stimulate memory
  - Can be useful to generate closed responses later
  - Useful at a hypothesis raising stage
- Inconvenient
  - Difficult to code and analyze
  - May be incomplete and / or unfocused

Examples:
- What disease can you acquire from tobacco?
- What places did you eat at in the week preceding the disease?

Closed questions: 1. Dichotomous options

- Suggested answers include 2 options only
  - Yes and No
  - Male and female
- Advantages
  - Forces a clear position
  - May be useful for key, important, well framed issues
- Inconvenient
  - May oversimplify issues

Examples:
- Did you eat at restaurant X between 1 and 28 February?
- Have you ever consumed tobacco products?
  - A dichotomous question here is likely to over-simplify, unless it is used as an introduction

Closed questions: 2. Multiple options

- Multiple options of answers are suggested
- One or multiple answers acceptable
- Advantage
  - Larger choice of answer options
- Inconvenient
  - May be difficult to choose only one option

Example:
- Where do you go to seek treatment for fever?
  - Government Hospital
  - Private clinic
  - Pharmacist
  - Traditional healer
- Do you wear a helmet when riding a bike?
  - Always
  - Sometimes
  - Never

Closed questions: 3. Quantitative answers

- The respondent must provide a quantified answer
- Advantage
  - Allows creation of continuous variables
- Inconvenient
  - May requires validation:
    - Some "quantified" answers might be looked in the way they can be handled as continuous variables

Examples:
- How many time did you visit the clinic in the last 12 months?
  - True continuous variable
  - Four visits is the double of two visits
- How would you describe your pain on a 1-10 scale where 1 would be the minimum and 10 would be the maximum?
  - In fact a qualitative variable with 10 options
  - Requires validation
    - It may not be the double of three on the scale

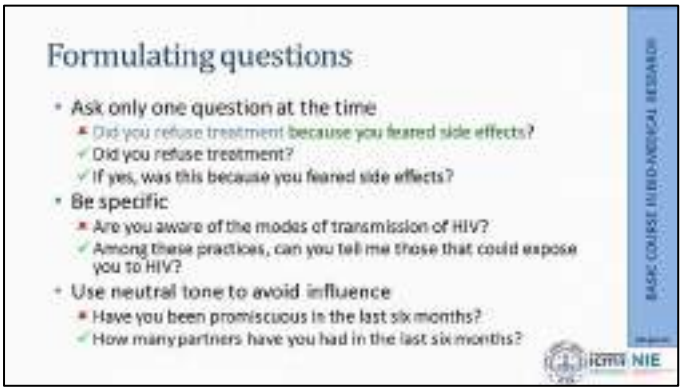
Semi-open questions

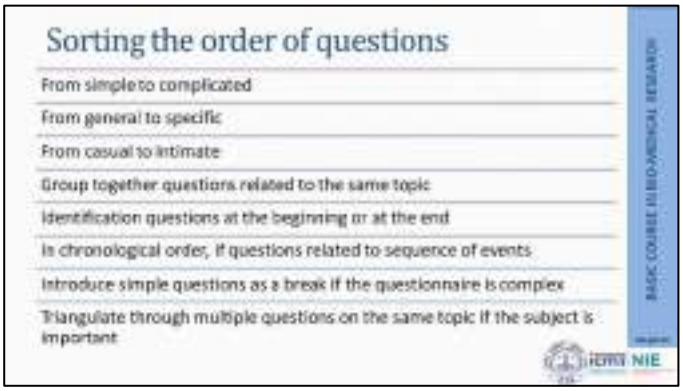
- Suggested answers
- Possibility to create another answer
  - Other, specify: _____
- Advantage
  - Leaves the door open to unplanned answers
- Inconvenient
  - Difficult to analyze

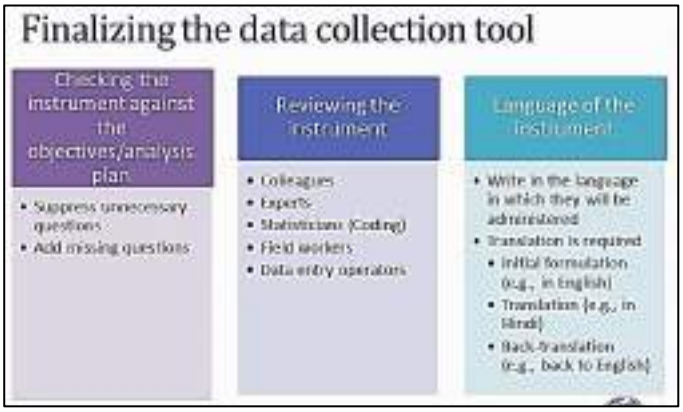
Example:
- Did your child have complication following measles?
  - None
  - Pneumonia
  - Diarrhoea
  - Eye problems
  - Other, specify: _____

Formulating questions

- Write short and precise questions
  - Full and complete phrases
  - Avoid ambiguities
- Use simple words of every day language
- Avoid negatives and double negatives
  - ✗ Do you sometimes care for patients without washing hands?
  - ✓ Do you systematically wash hands before caring for each patient?







1. Which component of the data collection instrument is constituted by open, closed and semi-open items?
   a) Introduction
   b) Identifier
   c) Questions
   d) Concluding statement

2. Self-administered questionnaire can be
   a) Paper-based or computer-assisted
   b) Used in face-to-face interviews
   c) Used in telephonic interviews
   d) All of the above

3. While formulating the questions, all the following need to be followed, EXCEPT
   a) Short and clear questions
   b) Avoid ambiguities
   c) Avoid words of every-day language
   d) Avoid negatives and double negatives

4. Structured observation guide
   a) Is useful to document certain processes
   b) Use checklist of items
   c) Can be used for in-depth interviews
   d) 'a' and 'b'

5. The interviewer does not provide options for responses in
   a) Open-ended questions
   b) Close-ended questions
   c) Semi-open questions
   d) All of the above

6. What is the disadvantage of closed questions with dichotomous options in a study questionnaire?
   a) Detailed information available
   b) Oversimplifies the issues
   c) Forces an unclear position
   d) May not be useful for key well framed issues

7. The information about participant's attitudes for behaviors such as wearing helmets, washing hands before eating, constitute
   a) Facts
   b) Knowledge
   c) Judgments
   d) Texts

8. The type of questions in which there is a possibility to add other answer in addition to the options suggested

a) Open questions
b) Semi-open questions
c) Closed questions
d) Close questions with multiple options

9. The type of questions in a questionnaire which allow creation of continuous variables as responses
a) Closed questions with quantitative answers
b) Open questions with quantitative answers
c) Both of the above
d) None of the above

10. All are true regarding the order of questions in a data collection tool EXCEPT
a) From intimate to casual
b) From general to specific
c) From simple to complicate
d) In chronological order, if questions related to sequence of events

11. A question was framed by an ophthalmologist as a part of data collection tool for her research-
"Which of the following symptoms you had in the last one week?" The options were – (1) Eye pain; (2) Redness of eye; (3) Watering of eye; (4) Low vision.
Given that a study participant may have multiple complaints, which of the following best describes the type of question?
a) Open question
b) Closed questions with dichotomous options
c) Closed question with multiple options
d) Closed question with quantitative answers

12. A researcher has framed a question in the research tool as- "What is the monthly income of the family?" This information constitutes-
a) Facts
b) Knowledge
c) Judgments
d) Healthy Life styles

13. Which of the following is correct in relation to an open question?
a) Answers are suggested
b) Stimulate memory
c) Easy to code and analysis
d) Freedom to respond is compromised

14. Glasgow coma Scale (GCS) is a scoring system to understand the consciousness level of a person. The score varies between 3 and 15. A researcher has included a question in research tool- "What is the GCS score during admission?"
This question is an example of-
a) Open question
b) Closed question with dichotomous option
c) Closed question with multiple option
d) Closed question with quantitative answers

15. An investigator wanted to study the clinical profile of patients presented with foreign body in nose, attended in the emergency department in the last 2 years in a hospital. Which of the following is the most suitable way to collect data?
a) Review of records
b) Cohort study
c) Randomized trial
d) Focus group discussion

16. Order of a question should be all, EXCEPT-
a) From simple to complicated
b) From general to specific
c) From intimate to casual
d) In chronological order

17. Structured observation guide
a) Is useful to document certain process
b) Uses checklist of items
c) Both 'a' and 'b'
d) None of the above

18. Which of the following factors related to data collection may lead to study failures?
a) Poorly defined research question
b) Vague timelines
c) Lack of supervision
d) All of the above

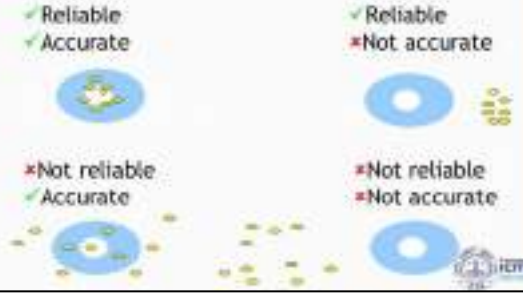
19. In which of the following type of question the interviewer does not provide options for responses?
a) Open-ended questions
b) Close-ended questions
c) Semi-open questions
d) Closed question with multiple options

20. 'Age in years' is commonly a continuous variable. However, a resident doctor decided to ask age as a closed question with dichotomous options (If age>65 years or <65 years). What is the disadvantage of such type of ques in a study questionnaire?
a) Detailed information available
b) Oversimplifies the issues
b) Forces an unclear position
d) It is easy to convert a dichotomous variable to a continuous variable

Data quality

- Reliability
  - Reproducibility/repeatability/precision
  - Ability of a measurement to give the same result or similar result with repeated measurements of the same thing
  - Refers to stability or consistency of information
- Accuracy
  - Ability of a measurement to be correct on the average

Reliability and accuracy



Six steps in data collection

1. Draft question-by-question guide
2. Train staff members who will collect data
3. Initiate data collection and ensure quality
4. Review collected data for quality and completeness
5. Debrief to trouble shoot difficulties
6. Validate

1. Draft question-by-question guide

- Short document to be understood as a guide for field workers
- Consider each question, number by number
- Provide guidance as to how the data should be collected
- Used as a road map for good data collection
  - Drafted initially
  - Revised as issues arise and are addressed

Example of Q by Q guide

- Question 6 (Housing):
  - Observe the house and note if made of mud or bricks
- Question 12 (Household income) :
  - Identify all the persons with financial income in the household
  - Estimate each source of income
  - Sum up to generate household income

2. Train staff who will collect data

- Select good, experienced investigators
- Present the study and its objectives
  - Slide presentation
- Distribute the q-by-q
- Walk people through the q-by-q
- List tasks to be conducted
- Answer questions
- Simulate interviews within the team

3. Initiate data collection and ensure quality

- Do pilot on site interviews under supervision
  - Note issues that may come up, resolve them as a group
  - Continue until the procedure is clear to everyone
- Plan data collection process with a supervisor and investigators
  - Ensure study forms are verified by the supervisor every day for any errors
  - Be available to answer questions
  - Do onsite visits
  - Do not press for quick completion

4. Review collected data for quality and completeness

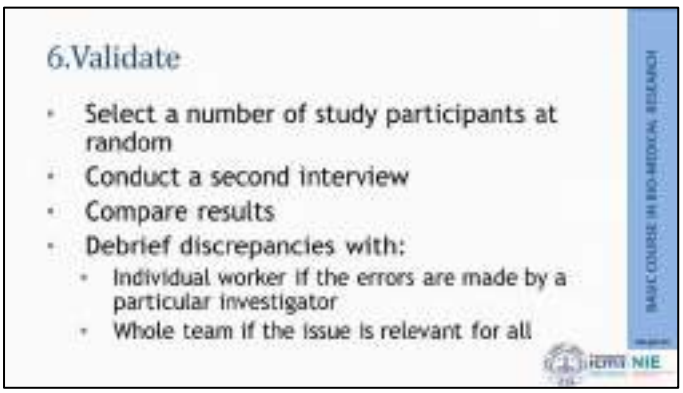
- Each team checks the data collection instruments before the respondent leaves
- The supervisor checks the instruments before leaving the location
- All take responsibility for the instrument:
  - Names and signatures
- Principal Investigator checks instruments as they come

Checks to conduct

- Completeness
  - Did the field worker fill all items?
- Readability
  - Is the writing readable?
- Consistency
  - Do the answer make sense?
  - Is there internal consistency?

5. Debrief to trouble shoot difficulties

- Regular meetings
  - Evening or morning
- Facilitate a discussion about
  - Issues identified
  - Clarification needed
- Make note of decisions on the q-by-q if needed



1. Reliability denotes
   a) Precision
   b) Repeatability
   c) Reproducibility
   d) All of the above

2. This should not be done in data collection
   a) Training of staff members
   b) Review of collected data for quality and completeness
   c) Manipulation of data
   d) Validation

3. Supportive supervision is essential for a good data collection process
   a) True
   b) False

4. The collected data should be
   a) Complete
   b) Readable
   c) Consistent
   d) All of the above

5. Which of the following is (are) true about the training of data collection staff?
   a) Conduct on-site training
   b) Conduct mock training sessions
   c) Training is always optional
   d) 'a' and 'b'

6. Which one of the following is the proper way of validating the data?
   a) Repetition of full data collection in the same population
   b) Data collection in new population
   c) Repetition of data collection in a randomly selected subset in the same population
   d) Repeat data collection is not required

7. Appropriate means to troubleshoot the difficulties in data collection process
   a) Regular review meetings
   b) Facilitate the discussion to identify issues during the review
   c) Clarify the issues experienced by staff during data collection
   d) All of the above

8. There is no need to present the study and its objectives to the field investigators
   a) True
   b) False

9. Which of the following statement is (are) true regarding data collection for an epidemiological study?
   a) Reliability refers to consistency of information
   b) Accuracy is the ability of a measurement to be correct on an average
   c) Feasibility is the ability of investigator to understand the data
   d) 'a' and 'b'

10. Time pressure during data collection may result in dilution of data quality
    a) True
    b) False

11. A data collection tool should be _____
    a) Valid
    b) Reliable
    c) Both 'a' and 'b'
    d) None

12. A neurosurgeon is planning for a hospital-based study on the patients coming to the emergency department with head injury. The collected data should be
    a) Complete
    b) Readable
    c) Consistent
    d) All of the above

13. Which of the following should not be done in relation to data collection?
    a) Training of staff members
    b) Review of collected data for quality and completeness
    c) Validation
    d) None of the above

14. State whether true or false: Piloting a data collection tool should be done under supervision

a) True
b) False

15. All of the following are true about a 'question by question guide', except-
    a) It is a document for the data collectors
    b) It helps in maintaining uniformity of the data collection
    c) It helps participants on how to respond
    d) It clarifies doubts on data collection

16. Time pressure during data collection may result in dilution of the data quality
    a) True
    b) False

17. Which one of the following is the proper way of validating the data?
    a) Repetition of full data collection in the same population
    b) Data collection in a new population
    c) Repetition of data collection in a randomly selected subset in the same population
    d) Repeat data collection not required

18. Which of the following statement is true regarding data collection for an epidemiological study?
    a) Reliability refers to consistency of information
    b) Accuracy is the ability of a measurement to be correct on an average
    c) Both 'a' and 'b'
    d) None of the above

19. Who is mainly responsible to check the accuracy of data collection instruments before leaving the location?
    a) Field investigator
    b) Field supervisor
    c) Principle investigator
    d) Study participant

20. All the following statements are true about training of the data collectors, except-
    a) Essential to ensure good quality data
    b) The investigators should choose the right people
    c) Communication skill is important for the data collectors
    d) Onsite training is not essential for data collectors

Data management includes

- Define variables
- Create study database and data dictionary
- Enter data and correct errors
- Create dataset for analysis
- Back up and archive the dataset

Key elements of data management

- Data structure
- Data entry
- Individual and aggregated databases
- Mother and daughter databases

Basic structure of a database

- Lines represent records
- Columns represent variables

	Identifier	Variable 1	Variable 2	Variable 3	Variable 4	Etc...
Record 1						
Record 2						
Record 3						
Etc...						

Data documentation

- Structure
  - Name, number of records etc
- Variables
  - Name, values, coding
- History
  - Creation, modification
- Storage information
  - Media, location, back up
- Additional information

Identifier in the database

- Unique
- Maintained by a computerized index
- Secured by quality assurance procedures

Using codes within the unique identifier

- Unique identifier may contain all information about that particular ID
- Each digit or set of digits refer to specific information
  - Example:
    - First and second digit: village
    - Third and fourth digit: Street
    - Fifth digit: House
    - Sixth and seventh digit: Person

Structure of the variables in the database

- Integer
  - Specify the number of digits
- Numeric
  - Specify the number of decimals
- Alpha-numeric
  - Specify length
  - Turn all letters to capitals
- Dates (specific format)

Creating variable names

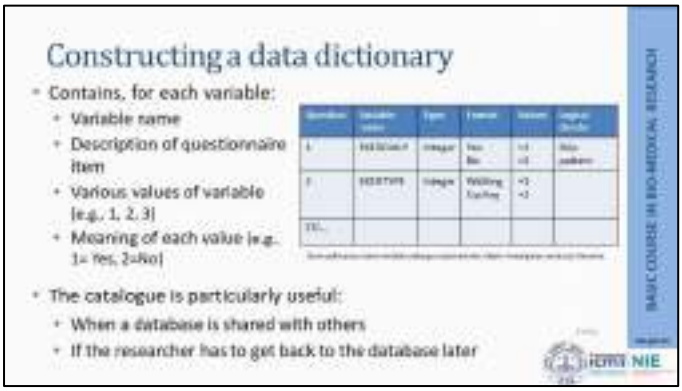
- Clear
  - Need to refer to the questionnaire item
  - Understandable (e.g., "EXERDAILY" for "Exercise daily")
- Short, no space
  - Most softwares require less than 10 characters
- Consistent
  - "EXERPAST" for "Exercise daily in the past"
  - "EXERCURRDLY" for "Exercise daily in the current"
  - "EXERPASTOCC" for "Exercise occasionally in the past"
  - "EXERCURROCC" for "Exercise occasionally in the current"
  - "VARIAB" for all crude variables (EXERCISE)
  - "VARIAB_12" for all dichotomized variables (EXERCISE_12)
- No duplicate
  - Trimming of names by software can create duplicate name

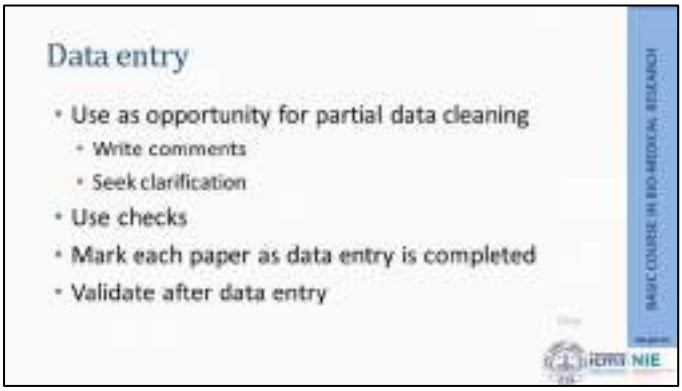
Design data entry-friendly data collection instrument

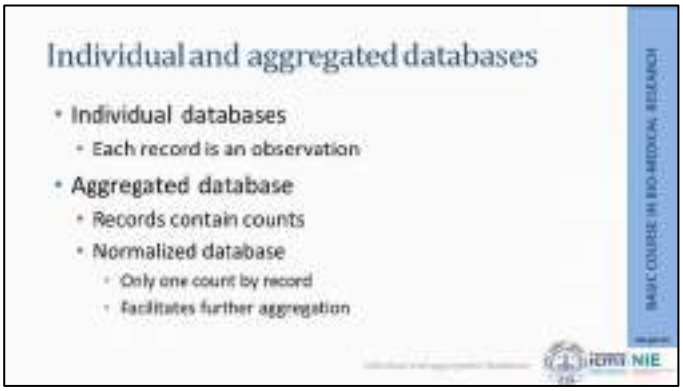
- Outline
  - Identifiers
  - Demographics
  - Outcome (Health problem/disease)
  - Exposures (variables, including third factors)
- Auto-coding function

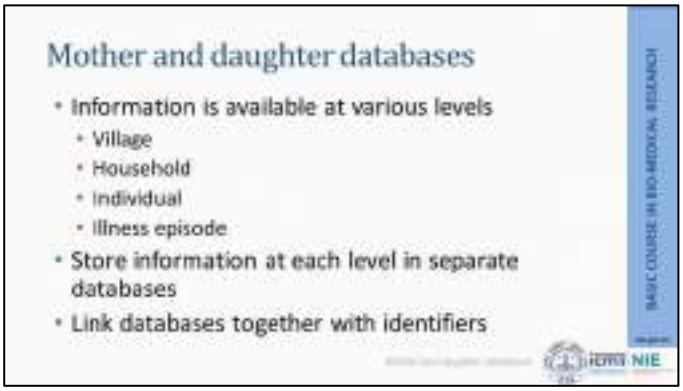
Coding

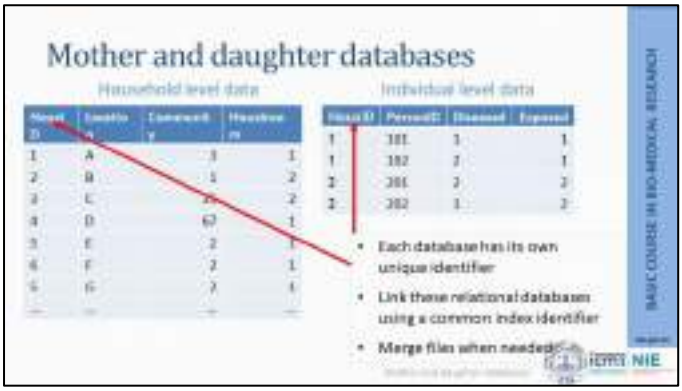
- Prefer numerical coding
- Decide on
  - Missing values (.) or (9, 99, 999)
  - Not applicable (8, 98, 998)
- Avoid cumbersome codes
  - WALKING (1) and CYCLING (2)
  - Doing WALKING and CYCLING (12)
- Use as "1" or "0" ("1" or "2") as baseline for gradients (Yes/No or Present/Absent) as appropriate depending on software for analysis

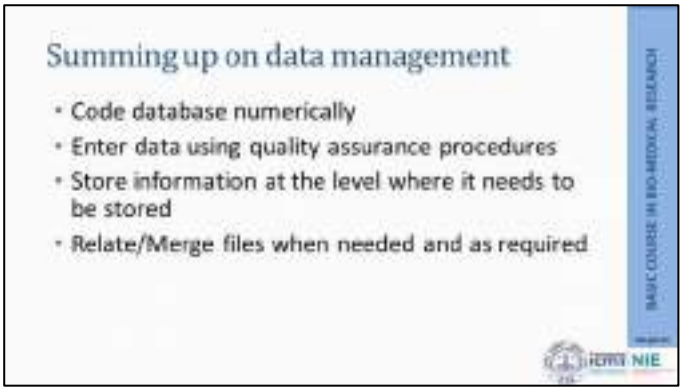












1. Steps in data management include
   a) Defining a variable, creating a study database and dictionary
   b) Enter data, correct errors and create data set for analysis
   c) Backup and archive data set
   d) All of the above

2. When we are creating variable name, it should be
   a) Clearly understandable and should refer to the questionnaire
   b) Long and can have spaces
   c) Consistent and without duplicates
   d) 'a' and 'c'

3. In a data management system, each row represents a
   a) Variable
   b) Record
   c) Heading
   d) Appendix

4. What is (are) the specifications that we need to check before doing data entry?
   a) Minimum and maximum values, legal codes, skip patterns
   b) Record name and description of record
   c) Automatic coding, coping data from preceding record and calculations
   d) 'a' and 'c'

5. Identifier in the database is (are)
   a) Unique
   b) Maintained by a computerized index
   c) Secured by quality assurance procedures
   d) All of the above

6. Key elements of data management
   a) Data structure and data entry
   b) Individual and aggregated databases
   c) Mother and daughter databases
   d) All of the above

7. The design of data collection instrument
   a) Data entry friendly
   b) Outline of major data collection topics/items
   c) Auto coding function
   d) All of the above

8. When we are coding for data entry, we should
   a) Prefer numerical coding
   b) Use highly complex codes
   c) Decide on the codes for 'missing values' and 'not applicable' items
   d) 'a' and 'c'

9. When information is available at various levels (e.g. at Village, Household, Individual and Illness episode), we can store information at each level in separate databases and link when necessary
   a) True
   b) False

10. Which of the following is (are) not true about normalized database?
    a) Normalized database facilitates further aggregation
    b) It has only one count by record
    c) Normalized database does not facilitate further data aggregation
    d) 'a' and 'b'

11. A post-graduate researcher has completed the data collection for her thesis. During data management, she should do all the following, EXCEPT-
    a) Applying for the ethics committee clearance
    b) Create study database
    c) Create dataset for analysis
    d) Back-up dataset

12. Data documentation includes information about the following items-
    a) Structure (Name, number of records etc) alone
    b) Storage information (Media, location, backup information)
    c) Structure (Name, number of records etc), Variables (Name, values, coding), History (Creation, modification), and Storage information (Media, location, backup information)
    d) Structure (Name, number of records etc), Storage information (Media, location, backup information), and Variables (Name, values, coding)

13. A variable name should be-
    a) Clearly understandable and should refer to the questionnaire
    b) Short, no space
    c) Consistent and without duplicates
    d) All of the above

14. Design of data entry can be broadly outlined as-
    a) Identifier, Demographics, Outcome, and Exposure
    b) Informed consent, Identifier, and Demographics
    c) Identifier, Demographics, Outcome and data analysis plan
    d) Informed consent, Identifier, Demographics, Outcome and data analysis plan

15. All of the following are true about 'Coding' a new variable, EXCEPT-
    a) Prefer numerical coding
    b) Decide on missing values while coding
    c) Avoid cumbersome codes
    d) coding with '0' and '1' should be avoided for dichotomous variables

16. Data entry can be considered as an opportunity to partially clean the data
    a) True
    b) False

17. A researcher in diabetes expected that that the fasting blood sugar levels may take any value between 50 and 150 gm/dL. In this research any coding of missing value as 99 may lead to an erroneous result.
    a) True
    b) False

18. While documenting the storage information of the database, we need to document
    a) Investigators information
    b) Time, place, person information
    c) Media, location and backup information
    d) Hardware configuration

19. Which of the following is incorrect in relation to the data catalogue?
    a) It describes all the variable for any future reference
    b) It is useful if we share the data with others
    c) It is useful to know how a variable has been coded
    d) It is advisable to exclude the missing values from data catalogue

20. Which of the following is incorrect about normalized database?
    a) Normalized database facilitates further aggregation
    b) It has only one count by record
    c) Normalized database does not facilitate further data aggregation
    d) Both 'b' and 'c'

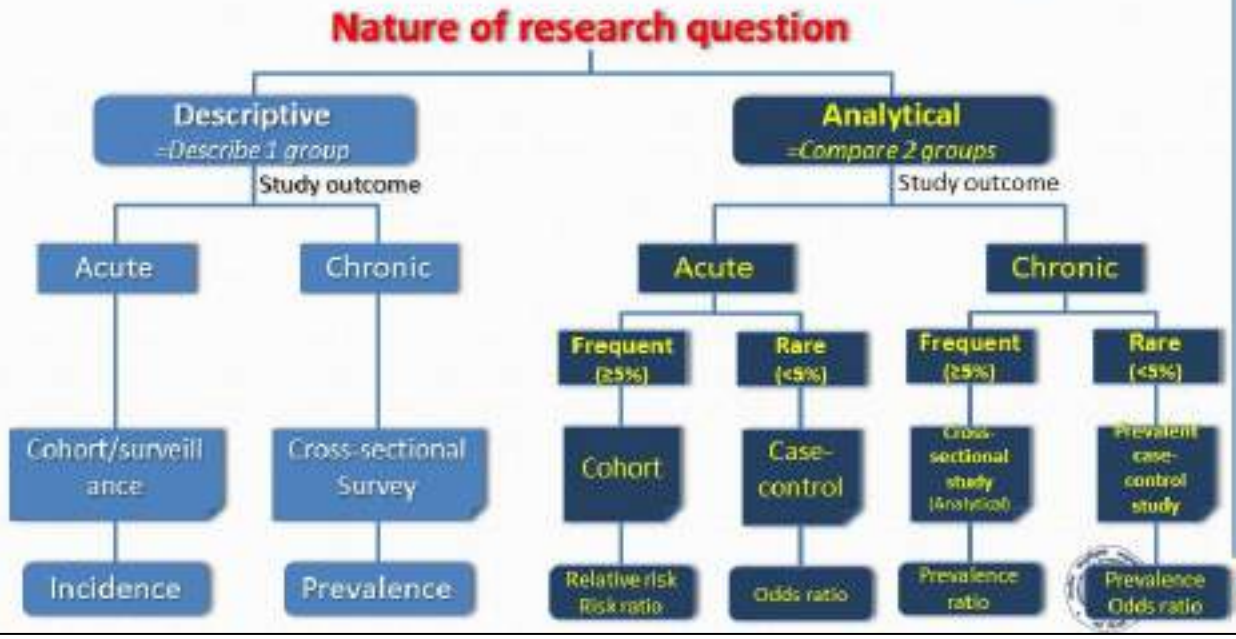
Key objectives of data analysis

- Plan the analysis
- Programme the crude analysis
- Deal with chance, biases and third factors
- Assess causality
- Measure clinical/public health impact

Sequence of data analysis strategy

1. Identify study type
2. Identify main variables
3. Become familiar with the data
4. Characterize study population
5. Examine outcome / exposure association
6. Create additional two-way tables
7. Conduct advanced analysis

Analysis plan depends on objectives and study type



BASIC COURSE IN BIO-MEDICAL RESEARCH

nie.gov.in

NIE

1. Identify study type

- Establish main analysis framework
  - Descriptive study (Estimation of a quantity)
  - Analytical study (Testing hypotheses)
- Get familiar with the study
  - Review protocol for study objectives and study type
  - Review questionnaire
  - Review analysis plan
  - Review data collection procedures
  - Obtain electronic database(s)
  - Decide on the software for analysis*

2. Identify main variables

- Outcomes
- Exposures
- Potential third factors
- Variables for subgroup analysis

3. Become familiar with the data

- Perform
  - Frequency distribution
    - Examine frequency of all the variables
  - Descriptive statistics
    - All the variables describing the study population
- Review number of observations by status in the database
  - Look for duplicates
  - Look for missing observations
- Check ranges and legal values
- Check consistency

4. Characterize study population

- Baseline characteristics
  - Distribution of study participants by socio-demographic- economic variables
    - e.g., Age, gender, income
- Frequency of clinical features/ health problems
- In analytical study → for compared groups

5. Examine outcome/exposure association

- Based on *a priori* hypotheses
  - Compare groups for frequency of exposures using appropriate measure of association
- Based on prior knowledge
- Based on study design

6. Create additional two-way tables

- Second-line analysis on the basis of findings
  - e.g., Creation of new variables

7. Conduct advanced analysis

- Dose-response
- Stratifications
- Multivariate analysis

Practical tips for data analysis

- Prepare data analysis in advance
  - Use empty table shells to prepare analysis
- Analyse by stages
  - Recoding
  - Descriptive
  - Analytical
- Avoid
  - *Post hoc* analysis
  - Data drenching

Initial stages of the analysis:

e.g., Effect of brisk walking on fasting blood sugar levels in diabetics

- Recoding stage
  - Create outcome data
  - Recode key variables e.g., age-groups, income
- Descriptive stage
  - Calculate frequency of outcome

Analytical stage of the analysis:

e.g., Effect of brisk walking on fasting blood sugar levels in diabetics

- Univariate analysis
  - Frequency of outcome by age, gender and income
  - Frequency of outcome by income categories (potentially examine dose-response effect)
- Stratified analysis
  - Frequency of outcome by income, stratified for age, gender and income
- Multivariate analysis
  - Logistic regression model

Software for data management and analysis

- ☒ Avoid spreadsheets for data management /analysis of any type /size
- ✓ Use software with data management & analysis tools
  - ✓ e.g., EpiInfo*

1. The three stages of data analysis are in the following order
   a) Descriptive stage, analytical stage and recoding stage
   b) Recoding stage, descriptive stage and analytical stage
   c) Analytical stage, descriptive stage and recoding stage
   d) Descriptive stage, coding stage, recoding stage

2. We need to avoid the following while performing data analysis
   a) *Post hoc* analysis
   b) Data drenching
   c) Stratified data analysis
   d) 'a' and 'b'

3. In the descriptive stage of analysis, we use logistic regression models
   a) True
   b) False

4. "Epi-Info" is a software used for data entry and data analysis
   a) True
   b) False

5. In analytical stage of data analysis, we perform the following in order
   a) Stratified analysis, univariate analysis and multivariate analysis
   b) Univariate analysis, stratified analysis and multivariate analysis
   c) Multivariate analysis, univariate analysis and stratified analysis
   d) Frequency analysis and univariate analysis

6. Among the seven steps of data analysis strategy, the sequence of data analysis is as follows
   (A) Conduct advanced analysis
   (B) Identify main variables
   (C) Become familiar with the data
   (D) Identify study type
   (E) Examine outcome/exposure association
   (F) Characterize study population
   (G) Create additional two-way tables
   a) A, B, C, D, E, F, G
   b) G, E, F, D, A, B, C
   c) D, B, C, F, E, G, A
   d) E, F, G, C, A, B, D

7. In case of descriptive studies, which of the following is wrong?
   a) We describe the study outcome for 1 group
   b) We compare the study outcome for 2 groups
   c) We calculate the incidence for cohort or surveillance data
   d) We calculate prevalence for cross sectional survey

8. If we are doing an analytical study and the study outcome is of acute nature and rare condition what is the appropriate (i) study design and (ii) measure of association?
   a) Cohort study - Relative risk
   b) Case-control study - Odds ratio
   c) Cross-sectional study - Prevalence ratio
   d) Surveillance - Incidence

9. Analysis plan depends on
   a) Objectives of the study
   b) Budget
   c) Study type (Descriptive or analytical)
   d) 'a' and 'c'

10. Use of spreadsheets, such as Excel, should be avoided for data management and analysis
    a) True
    b) False

11. Multivariate regression models are used during the descriptive stage of analysis
    a) True
    b) False

12. "Epi-Info" is a software that can be used to create data collection instrument format
    a) True
    b) False

13. In a research study the analysis plan depends on
    a) Objectives and study type
    b) Allocated budget
    c) Availability of the statistician
    d) Existing time for analysis

14. At the time of data cleaning, which of the following is not done?
    a) Checking and removing duplicates
    b) Dealing with missing observations
    c) Calculating strength of association
    d) Checking range and legal values

15. To describe the study population characteristics, we need to
    a) Calculate the frequency distribution
    b) Calculate measures of association
    c) Look for correlation between variables
    d) perform multivariable regression

16. While examining the association between exposure and outcome based on a priori hypotheses, we compare frequency of exposures between cases and controls using appropriate measure of association
    a) True
    b) False

17. If we are doing an analytical study and the study outcome is of acute nature and a frequent condition what is the appropriate (i) study design and (ii) measure of association?
    a) Cohort study - relative risk
    b) Case-control study – odds ratio
    c) Cross sectional study - Prevalence
    d) Surveillance - Incidence

18. Which of the following statements are CORRECT
    a) Plan for data analysis is made at the end of the study
    b) Recoding can be done for key variables
    c) Multivariate analysis is done before doing a univariate analysis
    d) Data drenching is acceptable

19. Which of the following is the correct sequence for data analysis
    i. Multivariate analysis
    ii. Recoding
    iii. Measures of association
    iv. Frequency distribution
    a) ii, iii, iv, i
    b) i, iii, iv ii
    c) ii, iv, iii, i
    d) iii, iv, i, ii

20. Spreadsheets are ideal tools for data entry and analysis
    a) True
    b) False

Ethical foundation is crucial for research, including health research

- Any research involving human participants should follow international standards of ethics
- Indian national standards are not less exacting and Indian ethical guidelines are on par with international guidelines
- Ethics review is also expected in situations involving no risk when available data are used or minimal risk such as when only questions are asked, no samples/ other specimens are collected

BASIC COURSE IN BIO-MEDICAL RESEARCH

Evolution of various guidance documents has greatly improved the practice of ethics in biomedical research

INTERNATIONAL	
1947, Nuremberg Code	Initiated discussion on rationale and justification of research[risk benefit analysis, competence of investigators and voluntary consent in research
1964, Helsinki Declaration, Revised 1883, 1989, 1996, 2000, 2008, 2013	Commitment to individual rights to make informed decisions, investigators' duties, research participants' welfare, vulnerability
1978-79, Belmont report	Described the basic ethics principles of autonomy, justice and beneficence, emphasized informed consent and review by ethics committee
1992-93, CIOMS guidelines [Council for International organizations on Medical Sciences and WHO], Revised 2002	Reporting of adverse drug reactions and safety of research participants, benefit-risk balance, need and principles of pharmacovigilance,
1996, ICH [International Council on Harmonization]	Good Clinical Practice

BASIC COURSE IN BIO-MEDICAL RESEARCH

NIE

Indian Council of Medical Research introduced Ethical Guidelines for Research on Human Participants

NATIONAL GUIDELINES	
National Ethical Guidelines for Biomedical and Health Research Involving Human Participants; ICMR, 2017. Available from https://www.icmr.nic.in/sites/default/files/guidelines/ICMR_Ethical_Guidelines_2017.pdf	All institutions in the country which carry out any form of biomedical research involving human beings should follow these guidelines in letter and spirit to protect safety and well being of all individuals.
There are several other national guidelines available	Genome Policy and Genetic Research [2000], Indian GCP [2001], Amendment of Drugs and Cosmetics Act [2002], Assisted Reproductive Technology [2005], Stem Cell Research and Bio-banking [2006]

BASIC COURSE IN BIO-MEDICAL RESEARCH

icmr NIE

Researchers must follow the following ethical principles

Respect for autonomy
- Autonomy is Latin for "self-rule" - We have an obligation to respect the decisions made by people concerning their own lives. This is respecting human dignity.
- We must not interfere with the decisions of competent adults, and also actively empower others for whom we are responsible.

Justice
- We have an obligation to provide all with whatever they are deserve. Basically, we have an obligation to treat all people equally, fairly, and impartially.
- All individuals should have an opportunity to participate in research unless contraindicated and we must not impose unfair burdens.

Beneficence
- We must be fair and correct in all our actions and must take positive steps to prevent harm.
- Sometimes in our endeavor to pursue beneficence, we place ourselves in direct conflict with respecting the autonomy of people.

Non-maleficence
- We must not harm others: "First, do no harm", and wherever harm cannot be avoided, we must try to minimize the same.
- It is wrong to waste resources that could be used for good.

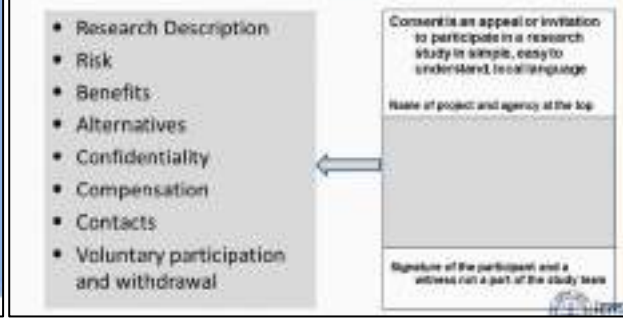
BASIC COURSE IN BIO-MEDICAL RESEARCH

What is informed consent?

- Informed consent is the process of informing the potential participants about the proposed research in a systematic manner and empower them to take an informed decision to participate in the research study
  - Understand study procedures and risks and benefits
  - Get all questions and concerns answered
  - Take a learned and informed decision to [or not to] participate
- This process can be repeated several times during the research study if necessary
- Although group consent is desirable [e.g. tribal studies], it cannot replace individual consent

Informed consent document

- Research Description
- Risk
- Benefits
- Alternatives
- Confidentiality
- Compensation
- Contacts
- Voluntary participation and withdrawal



Stakeholders in informed consent process

- Researchers and institutions:
  - Information – discussion and explanation – comprehension – voluntary decision
- Participants:
  - Informed, free and independent consent without coercion or force
- Sponsors, monitors, regulators:
  - Assess fairness of consenting procedure
  - Verify consent documentation of research participants

Issues related to informed consent

- **Whom does informed consent benefit?**
  - The research participant
  - The investigator
- **Is the research procedure adequately explained in the IC form?**
  - The language, simplicity and clarity
  - Translations and back translations, certification of translations
  - Test of understanding
- **Issue of witness to consent procedure**
  - Impartial witness
- **Can there be different types of informed consents**
  - Traditional written IC form
  - Pictorial consent

Importance of scientific review

Explores the scientific novelty, rationality and relevance

1. Justification for conducting the trial in the context of national priorities
2. Scientific merits of the research project and feasibility: Review of toxicological studies, laboratory and animal data
3. Technology transfer and capacity building at sites

Soundness of the study design:

Inclusion-exclusion criteria;	Sample size,
Randomization/ blinding procedures	End-point assessment
Study procedures and follow-up schedule	Pharmacy plan

Scientifically well planned research studies are more likely to correctly address human subjects and ethical issues

Objectives of regulatory review

- Evaluate pre-clinical trials data
- Assess in-country regulatory requirements for drug/ vaccine/ product import
- Ensure national requirements for special situations -genetically engineered products, stem cell research, research on reproductive technologies, organ transplantation etc.
- Sample shipment and transfers, transfer of raw data: IPR issues
- Exchange of scientists or visitors
- Budget: Foreign funding
- Research in border or high-security areas

Careful regulatory review results in answering some of the ethical concerns

Range of ethical issues that need to be addressed in health research

- Competence of the researchers and the research team
- Provisions for protection of human rights and ethical issues: vulnerable populations, women, children
- Measures for protecting confidentiality and non-discriminatory practices
- Appropriateness of informed consent and study specific educational material
- Mechanisms for reporting and management of adverse events and serious adverse events
- Care and support for research participants: standard of care, long-term care, post-trial access to care and product
- Reimbursement and compensation
- Continuing review of progress of the study

Main responsibility of institutional ethics committees or institutional review board

- Does the study have real/ potential individual/ community benefit?
- Are the rights of research participants adequately protected?
- Does the potential benefit far overweigh the risks associated with research participation?
- Will the participants and communities have access to study findings and benefits of research?
- What is the mechanism for provision of safety, care and support to research participants?

1. In which of the following guidelines, discussion on rationale and justification of risk benefit analysis of research and voluntary consent in research was initiated?
   a) Belmont report
   b) Helsinki declaration
   c) Nuremberg code
   d) CIOMS guidelines

2. In which type of the following study/studies is (are) informed consent not necessary?
   a) Investigation of an outbreak
   b) Analysis of mortality data of 2001-2010
   c) Using verbal autopsy to determine the cause of death
   d) All of the above

3. In which of the following type of research, ethical review is (are) mandatory?
   a) Prevalence of HIV infection using blood investigation
   b) Awareness about diabetes using questionnaire only
   c) Calculate out of pocket expenditure using secondary data
   d) All of the above

4. Which of the following is not important in the context of an informed consent document?
   a) Detailed description of study procedures
   b) Budget of the study
   c) Details regarding compensation and post trial access to care
   d) Contact details of the Principal Investigator

5. In the middle of a clinical trial, one participant decides to withdraw from the trial. But, the investigator pressurizes the participant to continue in the study till it completes. Which of the following ethical principles does the investigator violate?
   a) Justice
   b) Autonomy
   c) Beneficence
   d) Non-Maleficence

6. Which of the following are not ethical practices in health research?
   a) Taking informed consent from participants prior to study participation
   b) Giving lot of money to increase study participation
   c) Lack of adherence to study protocol
   d) 'b' and 'c'

7. "Do no harm" concept was emphasized in which of the following ethical principles?
   a) Justice
   b) Autonomy
   c) Beneficence
   d) Non-Maleficence

8. Genetic research that involves human participants and conducted by a private research institute should follow
   a) ICMR Guidelines
   b) Genome Policy and Genetic Research [2000]
   c) Both 'a' and 'b'
   d) Neither 'a' nor 'b'

9. While conducting research among tribal populations, which of the following is recommended?
   a) Consent from the tribal head (Group consent) is desirable
   b) Group consent can replace individual consent
   c) Women can be excluded from informed consent process
   d) Confidentiality not required

10. In a clinical research, the researcher knowingly excludes recruitment of female participants without any compelling indications. Which of the following ethical principle does the investigator violate?
    a) Justice
    b) Autonomy
    c) Beneficence
    d) Non-Maleficence

11. In which of the following situations is ethics review essential?
    a) When already available or archived data are used for research
    b) Involving some risk when some questions are asked, some samples are collected or some drugs are given

c) 'a' and 'b'
d) Only 'b'

12. A study participant can be forced by the investigator to continue in a trial against his will. This is in conflict with which of the following ethical principles?
a) Autonomy
b) Justice
c) Beneficence
d) None of the above

13. The physician should do what is medically indicated, do good than possible harm. This principle is encompassed in the ethical dimension of:
a) Beneficence
b) Justice
c) Nonmaleficence
d) Autonomy

14. Study monitors, regulators and ethics committee members have an authority to verify the consent documentation of research participants.
a) True
b) False

15. Which of the following is not true about an Informed Consent?
a) IC helps participants take an informed decision about participation in the research study
b) IC has information on potential risks and benefits of the study
c) IC process intends to protect the study participants
d) IC taken by coercion is considered valid

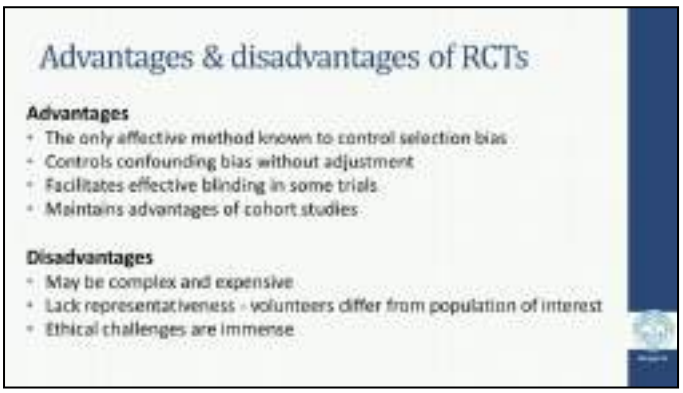
16. Which of the following have the guidance in the Belmont report?
a) The procedure of 'informed consent'
b) The basic ethics principles of autonomy, justice and beneficence
c) Review by ethics committee
d) All of the above
e) None of the above

17. Which of the following is typically not within the domain of ethical review of the proposed research?
a) Novelty of research
b) Competence of researchers
c) Relevance of research
d) To advocate for the study in the community

18. The process of Informed consent can be repeated several times during the research study if necessary.
a) True
b) False

19. As part of evaluation of a new vaccine which requires taking a daily oral dose of a refrigerated vaccine, the research team offers to provide a refrigerator to families of participants who don't have one. Which of the following ethical issues the Institutional Ethics Committee will have to deal with while reviewing the research study?
a) Undue inducement
b) Coercion
c) Compromising principle of justice
d) None of the above

20. In an observational study on menstrual hygiene among school going girls aged 14 to 16 years, informed assent will be required to be taken from the adolescent girls. In addition, informed consent will be required from:
a) Parent of adolescent girls
b) Institutional Head
c) Both
d) None of the above



1. A study design that randomly assigns participants into an experimental group or a control group is call as
   a) Cohort study
   b) Case-control study
   c) Randomized controlled trials
   d) Cross-sectional study

2. Which of the following statements is (are) true in case of adverse events in a clinical trial?
   a) An unexpected clinical/familial/social problem that occurs during treatment with a drug or other therapy is termed as adverse event
   b) Adverse events do not have to be caused by the drug or therapy under trial
   c) Temporal relationship between study product administration and adverse events is critically important
   d) All of the above

3. To ensure that safety and welfare of the research participants is adequately protected, it is important that the clinical trial protocol is critically reviewed for

the following
a) Scientific content
b) Ethical issues
c) Regulatory norms
d) All of the above

4. Informed consent is provided after explanation of
a) All study procedures
b) Risks
c) Benefits
d) All of the above

5. Bodies like Drug Controller General of India (DCGI) and Health Ministry Screening Committee (HMSC) are concerned with the following
a) Regulatory review
b) Scientific review
c) Ethics review
d) All of the above

6. Which of the following is (are) monitored in a clinical trial?
a) Adherence to Good Clinical Practice (GCP)
b) Documentation of informed consent, randomization and study product administration
c) Adverse events reporting
d) All of the above

7. The primary responsibilities of the Data Safety Monitoring Body (DSMB) are to
a) Periodically review and evaluate the accumulated study data for participant safety, study conduct and progress of trial
b) Periodically review and evaluate the accumulated study data for participant safety, study conduct and progress and make recommendation concerning the continuation, modification, or termination of the trial
c) Periodically make recommendations concerning the continuation, modification, or termination of the trial
d) Decide the randomization sequence

8. Which of the following best describes the advantages of conducting a Randomized controlled trial?
a) It is only effective design for overcoming selection bias of participants
b) The result can be readily generalized
c) It is a simple, uncomplicated and non-regulated study design
d) It requires small sample size

9. When is the Informed consent obtained from the subjects in a clinical trial?
a) Prior to participation in the trial
b) Just after the trial has started
c) At any point during the conduct of a trial
d) At the end of the trial

10. Investigators are required to report adverse events occurring during a clinical trial to which of the following agencies?
a) Regulatory authority
b) Sponsor
c) Institutional Ethics Committee
d) All of the above

11. A method of allocating treatment such that each subject has an equal chance of receiving any of the possible treatments in a clinical trial is known as:
a) Blinding
b) Randomization
c) Allocation concealment
d) None of the above.

12. Which of the following statements regarding document storage and archival after the conclusion of a trial; is correct?
a) If the data is computerized, there is no need to archive paper based records.
b) The investigator has a right to refuse to show the data even to regulatory authorities
c) Archival for a period of 5 - 15 years as per the requirement of the sponsor may be necessary
d) All of the above

13. Which of the following is not true?
a) Data Safety Monitoring Body (DSMB) is an independent entity.
b) DSMB is appointed by the Investigators
c) DSMB periodically reviews and evaluates the accumulated study data for participants' safety
d) DSMB assures that the scientific integrity of the trial is maintained during the period of interim analysis

14. An unexpected clinical/ familial/ social problem that occurs while on treatment with a drug or other therapy during participation in a clinical trial without any judgment about causality or relationship to the drug is known as:
a) Serious adverse event
b) Adverse event
c) Reportable event
d) None of the above

15. Clinical trials require review at various levels as per the in-country guidelines. State whether true or false.
a) True
b) False

16. The most common method of preventing potential harm to study participants is by adhering to 'trial stoppage rules' based on evidence on unacceptable toxicity or adverse effects rates seen during monitoring. State whether true or false.
a) True
b) False

17. Which of the following is true about screening protocol

of a clinical trial?
a) Those who are interested in participating in the trial participate in an interview may have to undergo medical examination
b) Eligibility of the potential participant is determined in screening
c) Information on study related procedures and inclusion and exclusion criteria are provided by the study investigators to the potential participants
d) All the above three statements are true
e) None of the above is true
f) Only 'a' is true

18. Reimbursements for which of the following raise no ethical questions?
a) Compensating for the time spent in coming over and the loss of daily wages due to participation
b) For the travel cost involved
c) For food expenses
d) Only 'a' and 'b'
e) 'a', 'b' and 'c'

19. Which of the following statements is wrong?
a) Drug Controller General of India is a Regulatory Authority in India
b) Institutional Governing Board is responsible for scientific review of projects
c) Institutional Ethics Committee is responsible for ethics review of a proposal
d) None of the above

20. Which of the following is NOT a method for identifying and preventing potential harm to study participants?
a) Adverse and serious adverse events reporting
b) Periodic review of the project by Data Safety Monitoring Board
c) Close watch on enrolment targets
d) Regular monitoring of the trial by a pre-identified monitoring agency

Competency to be gained from this lecture

- Write a concept paper for a research project

The seven steps of a successful protocol

1. Identify topic, question and objectives
2. Outline a one-page concept paper
3. Prepare dummy tables
4. Write draft protocol
5. Prepare instruments and annexes
6. Submit to peer review
7. Seek ethics committee review

The life cycle of research

The seven steps of a successful protocol

1. Identify topic, question and objectives
2. Outline a **one-page** concept paper
3. Prepare dummy tables
4. Write draft protocol
5. Prepare instruments and annexes
6. Submit to peer review
7. Seek ethics committee review

Rationale for using one-page concept paper

- Time is precious
  - For you
  - For your faculty / guide / reviewer
  - For funding agencies
- Brevity forces focus
- Many concept papers are not developed
  - Save time for an idea that may abort

Outline of the one-page, bullet-style, concept paper

1. Background and justification
2. Objectives
3. Methods
4. Expected benefits
5. Key references
6. Budget

Outline of the one-page, bullet-style, concept paper

Background and justification

- Importance of the problem
- What is known and unknown about the problem
- The information that is missing to address the problem effectively

Outline of the one-page, bullet-style, concept paper

Objectives

- 2-3 objectives
- Can be general and specific
- Can be primary and secondary

Outline of the one-page, bullet-style, concept paper

Methods

- Outline of the methods
- One bullet per point
  - Study design
  - Study population
  - Operational definitions
  - Sampling procedure
  - Sample size
  - Data collection
  - Analysis plan
  - Human participants protection

Outline of the one-page, bullet-style, concept paper

Key references

- Not more than 5
- As per standard guidelines
  - e.g., International Committee of Medical Journal Editors- *icmje.org*

Outline of the one-page, bullet-style, concept paper

Budget

- 4-5 lines
- No detailed justification
- Divided in salaries/per diem, travel, equipments & supplies and miscellaneous

Indian Council of Medical Research (ICMR), Department of Health Research, Govt. of India

ICMR's pre-proposal format

- Title of the project *(25 words)*
- Introduction *(250 words)*
- Novelty *(100 words)*
- Applicability *(100 words)*
- Description of the project *(700 words)*
  - Methodology, Feasibility, Outcome, Budget, etc

ICMR's Short-term studentship (STS) for medical undergraduates: **Format for STS proposals / project**

- Title *(25 words)*
- Introduction *(300 words)*
- Objectives *(100 words)*
- Methodology *(800 words)*
- Implications *(100 words)*
- References *(300 words)*

What can you achieve with one-page concept paper?

Patrick G Riley. The one-page proposal. 2002

Concept paper for observational studies – [insert title of the project here]

Insert name of primary investigator here

Background - justification (Do not use more than relevant references)

- Review in a few bullet the global consequences of the health problem being examined in terms of death, disability, effectiveness and cost-effectiveness of interventions. Avoid general statements and provide quantified data when available. Follow by explaining how this problem affects the place where the research is being considered.
- Provide information on completed, ongoing or planned prevention/control efforts targeting this problem in South Asia, India and / or the State/Place where the research will be conducted.
- Specify (1) the information needs of different stakeholders (health professionals/public health programme/policy makers & implementers) to improve this health problem (2) why currently available information is insufficient

Objectives

- Needs to be stated quantitatively for the primary outcome (Make it clear whether you propose to estimate a quantity or whether you propose to test a hypothesis)
- Clearly distinguish secondary from the primary objectives

- Clearly distinguish secondary from the primary objectives

Methods (Refer to www.equator-network.org for specific requirements for different studies)

Study population

- Specify the population to which you will undertake the study.

Study design

- Describe the type of study (e.g., survey, case-control, cohort studies) in one short bullet.

Operational definitions

- Provide information regarding the key definitions, criteria and / or control/recruitment strategy that you will be using.

Sampling procedure

- Describe the type of sampling you will be using.

Sample size

- Briefly mention your sample size and the main assumptions you used to calculate it. This should contain enough information for the reader to redo the calculations to check the estimate.

Data collection

- Explain shortly who will collect what kind of data, what the timeline is and what quality assurance mechanisms will be used.

Analysis plan

- Summarize the type of analysis (e.g., descriptive, analytical, stratified, multivariate) that you plan to carry out. Mention laboratory analysis if they will be part of the study.

Human participant protection

- Mention key measures taken to ensure the protection of human participants in your study and which ethics committee will review the proposal.

Expected benefits

- Describe the expected output (e.g., reports) that this study will generate and the timeline.
- Describe the expected outcome: How this study will influence management of this problem in question in the area where the research will be conducted.

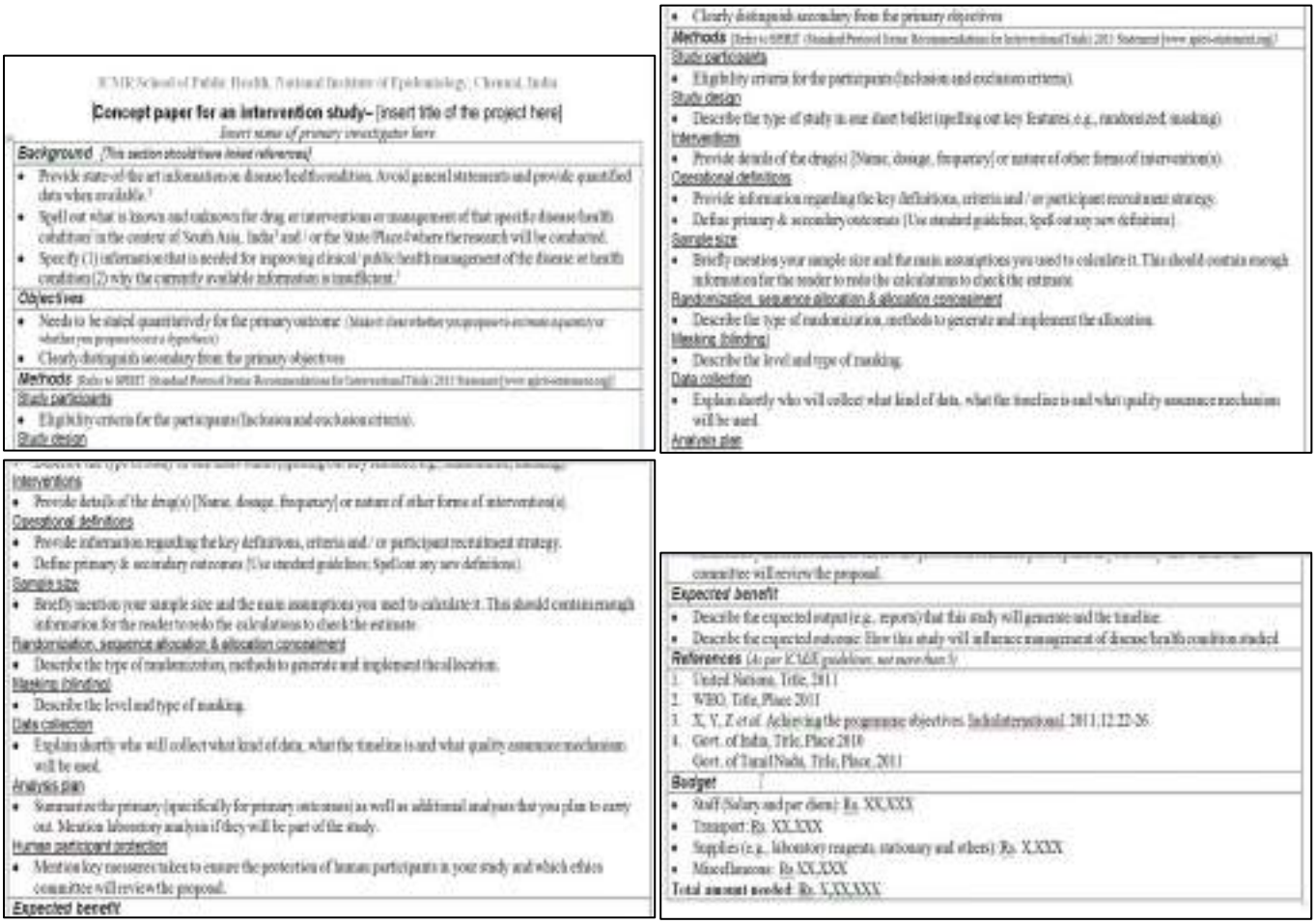
References (As per ICMJE guidelines, not more than 5)

1. United Nations, Title, 2011
2. WHO, Title, Place 2011
3. X, Y, Z et al. Achieving the programme objectives. India International. 2011;12:25-26
4. Govt. of India, Title, Place 2010
5. Govt. of Tamil Nadu, Title, Place, 2011

Budget

- Staff (Salary and per diem): Rs. XX,XXX
- Transport: Rs. XX,XXX
- Supplies (e.g., laboratory reagents, stationary and others): Rs. X,XXX
- Miscellaneous: Rs XX,XXX

Total amount needed: Rs. X,XX,XXX



Concept paper for an intervention study– [insert title of the project here]

Insert name of primary investigator here

Background *[This section should have limited references]*

- Provide state-of-the art information on disease load/condition. Avoid general statements and provide quantified data when available.[1]
- Spell out what is known and unknown for drug or intervention or management of that specific disease/health condition/ in the context of South Asia, India[2] and / or the State/Place[3] where the research will be conducted.
- Specify (1) information that is needed for improving clinical/ public health management of the disease or health condition; (2) why the currently available information is insufficient.[4]

Objectives

- Needs to be stated quantitatively for the primary outcome *[Insert clear criterion you propose to estimate (quantity) or whether you propose to test a hypothesis]*
- Clearly distinguish secondary from the primary objectives

Methods *[Refer to SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013 Statement [www.spirit-statement.org]]*

Study participants

- Eligibility criteria for the participants (inclusion and exclusion criteria).

Study design

- Describe the type of study in one short bullet (spelling out key features, e.g., randomized, masking)

Interventions

- Provide details of the drug(s) [Name, dosage, frequency] or nature of other forms of intervention(s)

Operational definitions

- Provide information regarding the key definitions, criteria and / or participant recruitment strategy.
- Define primary & secondary outcomes [Use standard guidelines; Spell out any new definitions]

Sample size

- Briefly mention your sample size and the main assumptions you used to calculate it. This should contain enough information for the reader to redo the calculations to check the estimate

Randomization, sequence allocation & allocation concealment

- Describe the type of randomization, methods to generate and implement the allocation.

Masking (blinding)

- Describe the level and type of masking.

Data collection

- Explain shortly who will collect what kind of data, what the timeline is and what quality assurance mechanisms will be used.

Analysis plan

- Summarize the primary [specifically for primary outcomes] as well as additional analyses that you plan to carry out. Mention laboratory analysis if they will be part of the study.

Human participant protection

- Mention key measures taken to ensure the protection of human participants in your study and which ethics committee will review the proposal.

Expected benefit

- Describe the expected output (e.g., reports) that this study will generate and the timeline.
- Describe the expected outcome: How this study will influence management of disease/health condition studied.

References *(As per ICMR guidelines, not more than 3)*

1. United Nations, Title, 2011
2. WHO, Title, Place 2011
3. X, Y, Z et al. Achieving the programme objectives. Indoleinternational. 2011;12:22-26
4. Govt. of India, Title, Place 2010
   Govt. of Tamil Nadu, Title, Place, 2011

Budget

- Staff (Salary and per diem): Rs. XX,XXX
- Transport: Rs. XX,XXX
- Supplies (e.g., laboratory reagents, stationary and others): Rs. X,XXX
- Miscellaneous: Rs XX,XXX

Total amount needed: Rs. X,XX,XXX

1. If there are many study objectives, it may be necessary to differentiate the objectives into primary and secondary or general and specific objectives
   a) True
   b) False

2. Which among the following is not a component of concept paper?
   a) Background and Justification
   b) Objectives and Methods
   c) Expected benefits, Key references and Budget
   d) Conclusion

3. Which of the following is NOT true about references in the concept paper?
   a) We can cite references in Introduction and Methods section
   b) It is important to write references following standard guidelines
   c) Statements should be linked to references
   d) We can have as many references as possible

4. The "Background and Justification" section in the concept paper should be written in the following sequence
   a) Known and unknown aspects of the problem, information that needs to be generated to address the problem in an effective manner and statement of objectives
   b) Known and unknown aspects of the problem, Importance of the study problem and information that needs to be generated to address the problem in an effective manner
   c) Importance of the study problem, known and unknown aspects of the problem and information that needs to be generated to address the problem in an effective manner
   d) Information that needs to be generated to address the problem in an effective manner, known and unknown aspects of the problem and statement of objectives

5. Advantages of writing a concept paper include
   a) You may be able to organize your ideas
   b) It gives an opportunity to stand out and receive a positive response from reviewers
   c) You are sure to get funding
   d) 'a' and 'b'

6. The elements of the methods section in the concept proposal needs to be adopted according to the study design chosen
   a) True
   b) False

7. The ethics section of the concept proposal should include information about
   a) Key measures taken to protect the study participants
   b) The ethics committee to which the study will be

submitted for approval
c) Scientific committee that will review the study
d) 'a' and 'b'

8. Which of the following needs to be spelt out in "Expected benefits" section of the concept proposal?
a) Expected outputs that the study will generate with timeline
b) Proposed immediate action based on research findings
c) How this research may set agenda for further research
d) All of the above

9. Budget estimate is not mandatory in the concept proposals. However, it would be very useful to prepare the indicative budget for key items
a) True
b) False

10. While writing the concept papers for intervention studies, the methods section must have the following details
a) Primary and secondary outcome definitions
b) Randomization, sequence allocation and allocation concealment
c) Dose, frequency, nature of Intervention
d) All of the above

11. Immediately after identification of research topic and statement of objectives, it may be preferable to
a) Write protocol
b) Outline one-page concept paper
c) Prepare dummy tables as per the analysis plan
d) Seek review by an institutional ethics committee

12. References need to be written following standard guidelines such as International Committee of Medical Journal Editors (ICMJE)
a) True
b) False

13. Which of the following is the guideline that can be used for drafting protocols for a clinical trial?
a) SPIRIT
b) PRISMA
c) CARE
d) STROBE

14. Why is concept paper necessary for a research project?
a) It helps to finish the data collection rapidly
b) It helps to organize the ideas
c) It helps to get instant approval of ethics committee
d) It helps to publish the research quickly

15. Which of the following components are included in the background and justification section of the concept paper
a) Context of the study problem
b) Operational definitions
c) Sampling technique
d) Study procedure

16. Which of the following information is not addressed in the ethics section of the concept paper?
a) Information about sample and data storage
b) Key measures taken to project the study participants
c) The ethics committee to which the study will be submitted for approval
d) Budget for salary of the projects staff

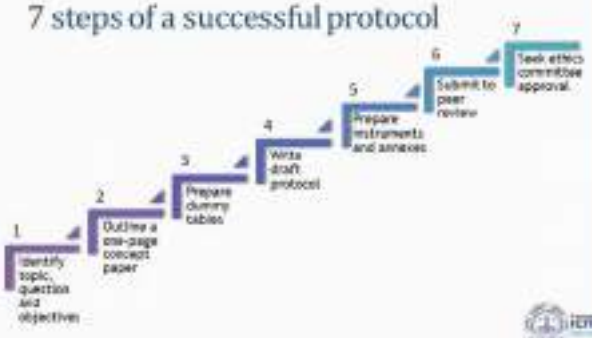
17. Which of the following is a component of concept paper?
a) Abstract
b) Objectives and Methods
c) Conclusion
d) Discussion

18. What is the basis for writing a one page concept paper?
a) Lack of time to draft a complete protocol
b) Overcomes inhibitions in drafting a complete protocol
c) It is mandatory for scientific committee protocol
d) For ethics committee approval

19. The indicative budget in a concept paper includes salaries, per diem, travel, equipment and supplies
a) True
b) False

20. The concept paper helps agencies for
a) Screening the proposal for funding
b) Scientific committee approval
c) Ethics committee approval
d) Publishing the manuscript

7 steps of a successful protocol



The first draft of the protocol

- Thought as it is written
- Keeps concept paper as summary
- Uses the concept paper outline
  - Background/justification
  - Objectives
  - Methods
  - Expected benefits
  - Budget
- Does not exceed 2000 words
  - Introduction < 20% of length
- Contains 5 – 10 key references

Outline of the methods section

- Study design
- Description of interventions (experimental studies)
- Study population
- Operational definitions
- Sampling procedure
- Sample size
- Data collection
- Analysis plan
- Project implementation plan (Quality assurance)
- Human subjects protection

Study design paragraph

- Explains how the objectives lead to indicators and to the study design
- Describes the type of study
  - Experimental
  - Cohort
  - Case control
  - Cross sectional
- Describes logistical arrangements
  - Prospective
  - Retrospective

Description of the interventions

- Applicable if an intervention is planned
  - Clinical trial
  - Community intervention
- Describes the "treatment" applied to the intervention and control group
  - Who?
  - What?
  - When?
  - How?

Study population paragraph

- Use time, place and person:
  - Inclusion criteria
  - Exclusion criteria
    - May be added as a separate section but do not differ conceptually from the inclusion criteria
- Do not confuse the study population and the study sample
- Ensure that the study population is suitable to address the objectives

Operational definitions paragraph

- Spells out and justifies
  - Key outcomes
  - Key exposures
- Clarity and specificity essential
- References, if applicable

Sampling procedure paragraph

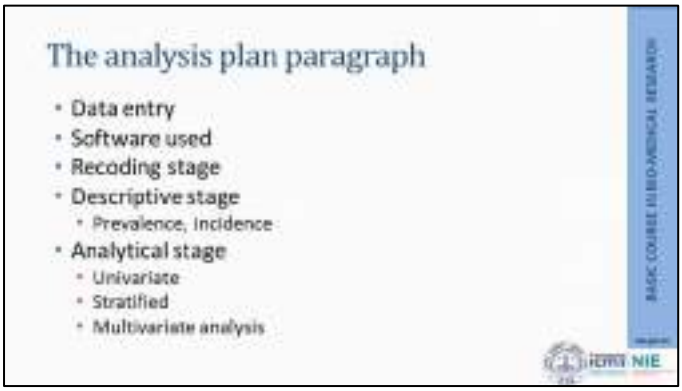
- Describes and justifies
  - The type of sample used
    - Convenience sample (Avoid if possible)
    - Random sample
    - Systematic sample
    - Cluster sample
  - The way the sample will be selected in practice
- Provides references, if needed
- Explains randomization, if applicable

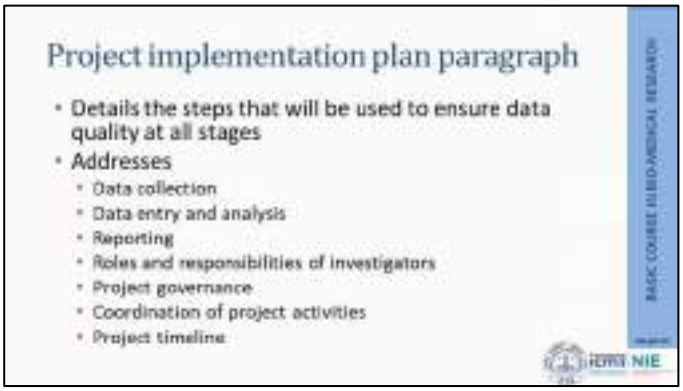
Sample size paragraph

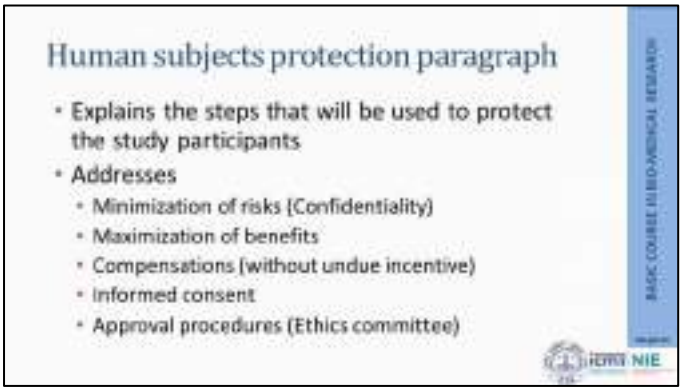
- Details all parameters used to estimate the sample size
- Explains how the estimate was generated
  - Software used
  - Formula used
- Provides references, if needed

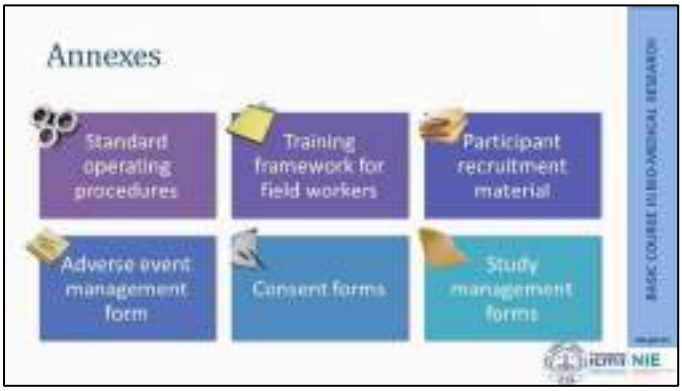
Data collection paragraph

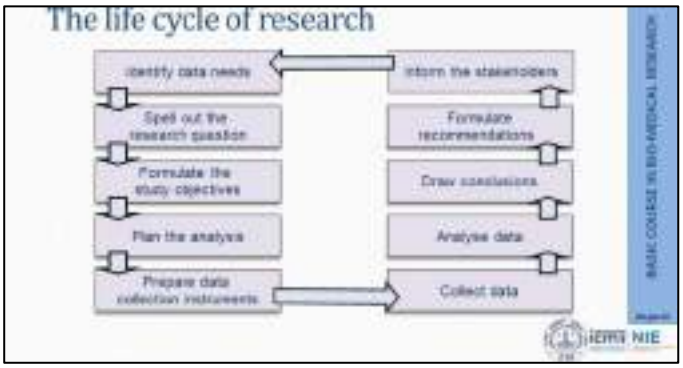
- Lists the data that will be collected
- Specifies how the data will be collected
  - Who?
  - How?
    - Type of instrument to be used
    - Type of data collection method











1. Description of the intervention is essential in a research protocol for
   a) Experimental study
   b) Case-control study
   c) Cohort study
   d) Cross-sectional study

2. Inclusion and exclusion criteria should be included under the following section in the protocol
   a) Sampling
   b) Study population
   c) Study design
   d) Sample size

3. The details regarding data quality assurance should be written in the following section in the protocol
   a) Data collection
   b) Data analysis
   c) Project implementation
   d) Data entry

4. The following annexure in the study protocol deals with toxicity management
   a) Study management forms
   b) Standard operating procedures
   c) Consent forms
   d) Adverse event management form

5. Study population, sample size and sampling are included in the following section of the protocol
   a) Introduction
   b) Methods
   c) Objectives
   d) Expected Benefits

6. First step for writing a successful protocol
   a) Write a one page concept paper
   b) Identify topic, research question and objectives
   c) Write a draft protocol
   d) Seek ethics approval

7. Key outcomes and exposures should be explained under
   a) Data analysis
   b) Data collection tools
   c) Sampling
   d) Operational definitions

8. Human participant protection paragraph addresses all except
   a) Confidentiality
   b) Risks
   c) Compensation
   d) Sample size calculation

9. Willingness of 'study participants' to participate in the study is obtained by
   a) Informed consent
   b) Oral commitment
   c) Willingness not necessary
   d) None of the above

10. The section that guides how the objectives lead to indicators
    a) Introduction
    b) Study design
    c) Budget
    d) Objectives

11. Which of the following statements is True regarding the first draft of the protocol
    a) The draft is the final document and has to be adhered to as it is.
    b) The concept paper can be used as an outline for drafting the first draft of the protocol.
    c) Background with justification, method of conducting the study and expected benefits are stated briefly as in the concept paper.
    d) Additional references must not be added.

12. It is ideal that the first draft of the protocol
    a) Exceeds >2000 words
    b) Does not exceed >2000 words
    c) Exceeds >3000 words
    d) Does not exceed >3000 words

13. Data collection paragraph in the protocol should specify all, EXCEPT
    a) The kind of data that will be collected
    b) Information about the data collector involved in data collection
    c) The detailed manner in which the data collector is going to collect the data
    d) The details of how the collected data will be used for policy recommendations

14. Mode of data entry, software for data analysis and plan for data analysis are included in the following section of the protocol
    a) Introduction
    b) Methods
    c) Objectives
    d) Expected Benefits

15. Which of the following is a step for drafting a successful protocol?
    a) Writing an abstract
    b) Submitting for peer review
    c) Seeking consent from participants
    d) Data analysis

16. Which of the following statements is true?
    a) It is sufficient to mention whether the study is quantitative or qualitative in study design section
    b) The concept paper can contain more than 20 references relating to the study
    c) Sample size calculation is not necessary for conducting research
    d) Human subject protection statement should be included in the methods section

17. The introduction can be 40% of the content of the protocol
    a) True
    b) False

18. Which of the following sections mentions about the detailed plan for conducting the study
    a) Introduction
    b) Results
    c) Methods
    d) Discussion

19. Which of the following is a part of introduction section of the protocol?
    a) Inclusion and Exclusion criteria
    b) Detailed budget
    c) Participant safety and protection
    d) Background with justification

20. Informed consent, procedures for minimizing participant risk and compensations are included in which of the following sections
    a) Introduction
    b) Abstract
    c) Human subject protection
    d) Study procedure

Learning objectives

At the end of the session, the participants will be able to-

1. Recognize various ethical issues related to publication
2. Make use of the guidelines available from various national and international organizations for publication ethics

Recap: Life cycle of a research

Why to publish a research finding?

Promotes academic/ career progression

(Medical Council of India, 2017)

Allows us to communicate research findings

Identifies research gaps, potential areas of future research

Increases responsibility to influence practice

Postgraduate biomedical research in India

- Lack of relevance of research question
- Not addressing the local needs
- Inappropriate design and methods
- Inaccessibility of the full report

Postgraduate research contributes to clinical practice

Components of Publication Ethics

1. Ethics review/ Breach of confidentiality
2. Fabrication and Falsification
3. Authorship
4. Plagiarism
5. Ethics related to submission
6. Conflict of interest

1. Ethics review/ Breach of confidentiality

Human/ animal ethics committee approval as per National guidelines

Informed consent/ assent- A must when you are conducting research with human participants

Trials- Registration with 'Clinical Trial Registry of India' (CTRI)

Data confidentiality- Without institutional permission, you can not share your data with someone for analysis who is working in a different institution

2. Fabrication/ Falsification

- If the research results not generated from the study (fabrication) or generated by manipulating data (falsification)
- Extremely serious misconduct
- All the case report forms and the data should be preserved

Example of falsification and fabrication (1/2)

A final year postgraduate student came to me requesting to analyse his thesis data. He also asked if I could make some significant results, though they weren't initially so!

Falsification

Curiously I asked how he has collected the data. To my surprise, he revealed that the data was taken from his senior's thesis.

Fabrication

Example of fabrication and falsification (2/2)

Kyoto University Finds Stem Cell Researcher Guilty of Data Fabrication

The scientist was a member of a stem cell research team led by Nobel laureate Shinya Yamanaka.

Breech of trust of the common people/ loss of time and the resources

3. Authorship

- Authorship confers credit, implies responsibility and accountability of the published work
- Each author should be clear about their responsibility
- It is mandatory to declare the contribution of each author
- Decide on authorship while writing protocol
- The ICMJE (International Committee of Medical Journal Editors) recommends that authorship be based on 4 criteria

Criteria for authorship

1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
2. Drafting the work or revising it critically for important intellectual content; AND
3. Final approval of the version to be published; AND
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved

4. Ethics related to submission

- Simultaneous submission- submitting the same manuscript simultaneously at the different journals
- Duplicate publication- Submitting a new manuscript with same hypothesis, methods, data, discussion and conclusion
- Self-citation- citing own publication out of context
- Predatory journals- Hardly peer reviews any manuscript

List of predatory journals- University Grants Commission-Consortium for Academic and Research Ethics (UGC-CARE)

5. Plagiarism

- Use of previously published manuscript by someone for his/ her manuscript or unreferenced use of other's published and unpublished ideas without consent, credit, or acknowledgement
- Most common form of plagiarism is copying text word-for-word
- Considered as serious publication misconduct

Types of Plagiarism

- Direct Plagiarism
  - Complete or partial copying without acknowledging the original author
- Self-Plagiarism
  - Duplicates of previous works or sentences
- Redundant publications ('salami' publications)
  - Publishing similar manuscripts/reports based on the same experiments

How to avoid plagiarism?

- Avoid "copy-paste"
- Write the concept in own words- spend more time
- Acknowledge original sources (Even unpublished works)
- Cite references accurately
- Avoid writing several articles of the same type
- Use anti-plagiarism software tools like *'URKUND'*, *'Ithenticate'* or *'Turnitin'* etc. (as per University norms)

6. Conflict of interest (COI)

- Financial, personal, social or other interests that directly or indirectly influence the conduct of the author with respect to the manuscript
- Example- A PG/ researcher is conducting a drug trial which is funded by a pharmaceutical company.
- Disclose any such conflicts during submission (Mandatory)
- Readers can determine the influence of such COI on conclusion of the paper

Consequences of research misconduct

COPE (Committee on Publication Ethics) describes the consequences:

- Depends on the type of misconduct (Major/ Minor)
- Author can get blacklisted by member journals
- Institute can take action

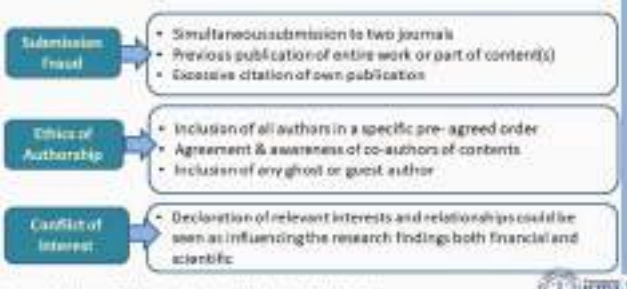
Publication criteria for- "Minimum Qualifications for Teachers in Medical Institutions (Amendment) Regulations, 2019"

Publication specifics	Guidelines
Index agencies	Should be published in journals included in Medline, Pubmed Central, Citation index, Sciences Citation index, Expanded Embase, Scopus, Directory of Open access journals (DoAJ) will be considered)
Type of articles	original papers, meta-analysis, systematic reviews, and case series
Authorship	The author must be amongst **first three** or should be the Corresponding author

The Gazette of India
BOARD OF GOVERNORS IN SUPERSESSION OF MEDICAL COUNCIL OF INDIA, New Delhi, the 12th February, 2020

Checklist for Publication Ethics

Checklist for Publication Ethics



1. All of the following statements regarding research publications are correct, Except
   a) Publishing paper is important for getting promotion in academic institutions
   b) Publishing research findings help to identify the research gaps
   c) Negative findings in a research should not be published
   d) Publishing research findings improves the credibility of a researcher

2. A senior resident of Psychiatry department of a medical college wrote a manuscript based on his thesis work. He has put his wife's name as a co-author who is working in the Physiology department of the same college. Which of the following statements supports the act of the senior resident in providing authorship to his wife?
   a) He can give authorship to anyone since it is his research work
   b) The guide should decide on who should be the authors
   c) His wife has contributed in designing the residents' thesis work
   d) It is not a good practice to include researcher from different department as authors

3. Which of the following is incorrect about authorship?
   a) Authorship confers credit, implies responsibility and accountability of the published work
   b) International Committee of Medical Journal Editors recommends criteria on authorship
   c) It is mandatory to declare the contribution of each author
   d) It is not mandatory that all authors should approve the final version of the manuscript

4. Which of the following is false about plagiarism?
   a) It can be copying and pasting of contents from a published manuscript
   b) It can be copying someone's idea
   c) It is not considered as a serious publication misconduct
   d) *'Urkund'* is one of the software used to check plagiarism

5. Which of the following is (are) the consequence(s) of plagiarism of manuscript?
   a) The journal can retract the manuscript
   b) Institute can take action on the author/researcher
   c) The researcher loses professional reputation
   d) All of the above

6. You have finished writing a manuscript and plan to publish it. Which of the following is the best practice?
   a) Submit to multiple journals at the same time
   b) Submit to a journal and wait for the journal's response
   c) Submit to many journals; once it gets published in one journal, withdraw it from the other journals
   d) Submit the same manuscript in different languages to different journals

7. A researcher conducted a study to identify risk factors for exacerbation of bronchial asthma. The researcher was due for job promotion. However, the researcher was lacking enough publications to ensure promotion. Hence, in order to have maximum number of publications from the work, the researcher decided to produce three different manuscripts instead of one manuscript comprehensively covering all aspects of the study. What is this act called?
   a) Plagiarism
   b) Falsification
   c) Salami slicing
   d) Fabrication

8. Which of the following organizations directly deals with publication ethics?
   a) Indian Medical Association (IMA)
   b) Committee on Publication Ethics (COPE)
   c) World Health Organization (WHO)
   d) Joint National Committee (JNC)

9. A group of researchers submitted a manuscript for publication based on a drug trial. Because they did not register under the clinical trial registry of India (CTRI), one reputed journal rejected the paper. The researcher resubmitted the paper in a different journal and this journal published it without asking any queries.
   Which of the following is the correct statement?
   a) It is necessary to register all drug trials under CTRI
   b) The journal which published the paper is likely to be a predatory journal
   c) Both 'a' and 'b' are correct
   d) None of the above

10. Among the following which is the best practice for determining the authorship?
    a) Authorship can be based on the criteria given by ICMJE
    b) Authorship should be decided after submission to a journal
    c) It is necessary to include head of the department/institution as a co-author
    d) Authorship can be gifted to friends even if they have not contributed to that study

11. Which of the following is incorrect about publishing a research work?
    a) Publishing paper is important for getting promotion in academic institutions
    b) Publishing research findings helps to identify the research gaps
    c) Common people should not read such research

findings
d) None of the above

12. All clinical trials in India should be registered with Clinical Trial Registry of India.
a) True
b) False

13. A neonatologist planned to conduct a clinical trial to explore the effect of intervention X on hypothermia of the newborn children (Age <7 days) over intervention A (The current practice). All the following are true about the trial, EXCEPT-
a) Ethics Committee approval is a must to conduct the trial
b) The trial should be registered under the Clinical Trial Registry of India
c) Informed consent should be taken from either of the parents
d) Age appropriate assent is a must in this trial

14. Manipulating data is known as fabrication
a) True
b) False

15. Which of the following is correct about determining the authorship?
a) The investigators can follow ICMJE guideline to determine authorship
b) The sequence should always be based on alphabetical orders
c) The investigators should include the head of the institution's name irrespective of his/ her contribution
d) None of the above

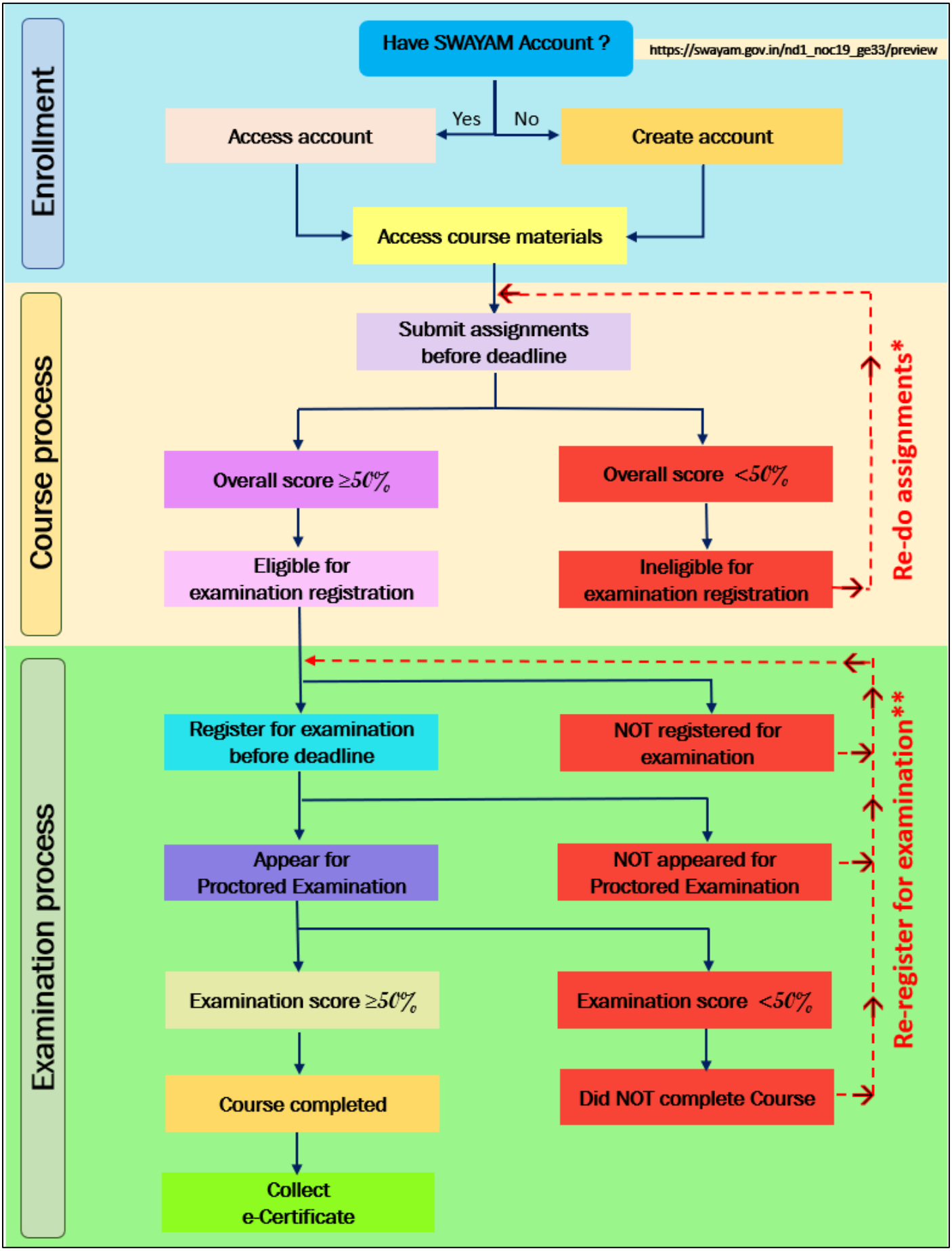
16. A group of researchers submitted a manuscript in a reputed journal. Even after 5 months of submission, they did not receive the peer review comments from the journal. The authors decided to submit the manuscript to a different journal without informing the editor of the previous journal. Which of the following term describes the situation best?
a) Duplicate publication
b) Simultaneous submission
c) Self-citation
d) Breach of confidentiality

17. Dr. D has copied the idea of Dr. A for his thesis. Copying an idea shouldn't be considered as plagiarism.
a) True
b) False

18. Dr. A is in the process of writing review of literature for her thesis. Her guide has instructed her to avoid plagiarism. Dr. A should take all the following measures to avoid plagiarism, EXCEPT-
a) Avoid copying and pasting
b) Acknowledge original sources
c) Take help of anti-plagiarism software
d) Copy from her own previous work

19. Which of the following is correct in relation to the conflict of interest?
a) Conflict of interest is always financial
b) Conflict of interest necessarily changes the outcome of interest
c) It is recommended to hide the COI during submission of a manuscript
d) Readers can determine the influence of COI on conclusion of the paper

20. An editor of a reputed journal found that most of the finding of a manuscript matches with a previously published paper by different authors. The editor considered it as a case of plagiarism. Which of the following about plagiarism is true?
a) The journal can retract the article, if already published
b) The editor can inform the authors' institute about it
c) The researchers may lose their reputation
d) All of the above


Enrollment
Have SWAYAM Account ?
https://swayam.gov.in/nd1_noc19_ge33/preview
Yes
No
Access account
Create account
Access course materials
Course process
Submit assignments before deadline
Overall score ≥50%
Overall score <50%
Eligible for examination registration
Ineligible for examination registration
Re-do assignments*
Examination process
Register for examination before deadline
NOT registered for examination
Appear for Proctored Examination
NOT appeared for Proctored Examination
Examination score ≥50%
Examination score <50%
Course completed
Did NOT complete Course
Collect e-Certificate
Re-register for examination**