Data Analysis in Bio Medical Research
Determination of sample size and Power analysis - Statistical implications on Bio Medical Research

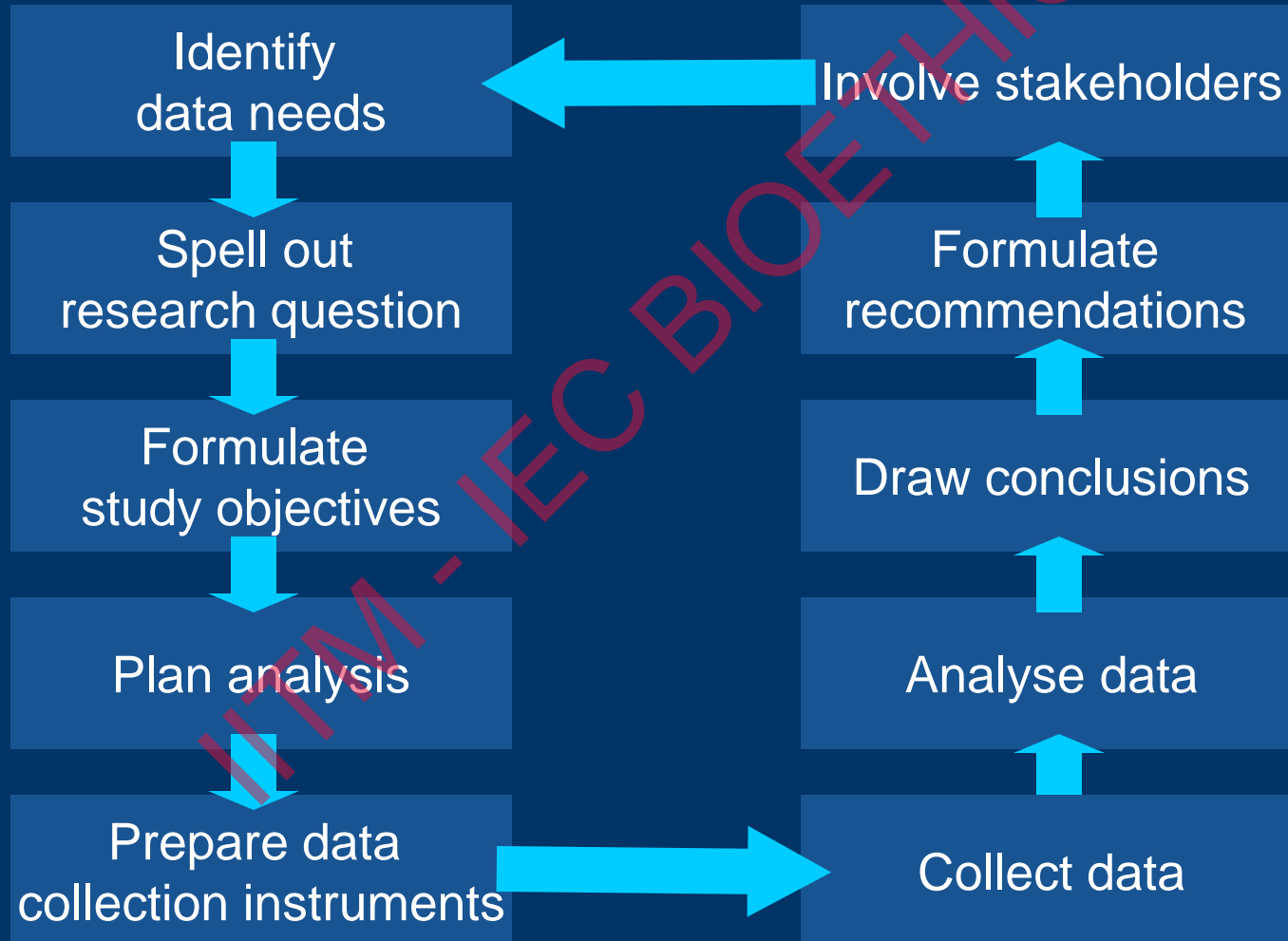Dr. R. Ramakrishnan

Ex-Scientist G

ICMR-National Institute of Epidemiology

Chennai, Tamil Nadu

IIT, Madras, 22 June 2023

The life cycle of research

Key areas to be considered

- Objectives of study

- Design and indicators

- Study parameters

- Analysis

- Sample size

The ad-hoc approach to conducting research

Before data collection

- I want to do a study
  - I am not clear about objectives
- I prepare a questionnaire
  - I am not clear about what information I need
- I collect data
  - I am not clear what I will use for what

After data collection

- I come back with data
  - I realize they are difficult to analyse
- I analyse the data
  - I realize it is difficult to interpret results
- I interpret the results
  - I realize it is difficult to use them

Sounds familiar?

A road map to making sense of data

1. Formulate the objectives of the study
2. Choose a design to identify key indicators
3. Identify parameters needed for indicators
4. Prepare the analysis
5. Estimate sample size

Objectives

The analysis plan:
A road map to making sense of data

1. Formulate the objectives of the study
2. Choose a design to identify key indicators
3. Identify parameters needed for indicators
4. Prepare the analysis
5. Estimate sample size

Objectives

The study objectives

- Formulated in limited number

- Sorted out as primary and secondary

- Focused

- No more than one verb each

- Clear about whether:

  - Estimating quantity **<Descriptive>**

  - Testing hypothesis (eg.) **<Analytical>**

- Epidemiological terms

Objectives

Estimating versus testing

- Estimating a quantity **\<Descriptive\>**
  - Use the verb "Estimate"
    - e.g., Estimate the prevalence of depression
- Testing a hypothesis **\<Analytical\>**
  - Use the verb "Determine"
    - e.g., Determine whether stress caused suicides

Objectives

Good and bad examples of study objectives

- Determine importance of hypertension

  - Estimate prevalence of hypertension in the community

- Assess yoga practice and hypertension

  - Estimate effect of yoga practice on hypertension control

- Evaluate sleep disturbance and depression

  - Determine whether depression is more common among those with sleep disturbance

Objectives

The analysis plan:
A road map to making sense of data

1. Formulate the objectives of the study
2. **Choose a design to identify key indicators**
3. Identify parameters needed for indicators
4. Prepare the analysis
5. Estimate sample size

Design and indicators

Elements to consider to choose study design

1. **Is the study descriptive or analytical?**
   - Is there a need to compare groups?
   - Is there just a need to estimate a frequency?
2. **Is the outcome (e.g., disease) acute or chronic**
   - Need of prevalence data for chronic outcomes
   - Need of incidence data for acute outcomes
3. **Is the outcome common or rare?**
   - Case-control design for rare outcomes
   - Cohort / cross-sectional designs for common outcomes

Design and indicators

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

National Institute of Epidemiology Chennai

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"

National Institute of
Epidemiology Chennai

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.

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

National Institute of Epidemiology Chennai

SAMPLE SIZE : Free Soft wares for Sample Size

**OpenEpi
Supported by Centers for Disease Control and Prevention, Atlanta
www.openepi.com**

**PS: Power and Sample Size Calculation
by Department of Bio statistics
Vanderbilt University**
http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize

National Institute of
Epidemiology Chennai

The analysis plan determines the sample size

- Choose the study design
  - Cohort, case-control or survey
- Determine the level
  - Descriptive or analytical
- Common mistake:
  - Designing a descriptive study
  - Trying comparisons for which sample size is insufficient

Sample size

Summing up

- Research protocols require clear analysis plan
  - Precise, focused objectives
  - Plan for design, conduct and analysis
- Sample size very much depends on the design and objectives.
- Write draft, peer-review, revise, revise, finalize protocol

Thank you