



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

Overview of data analysis

Dr. Manickam Ponnaiah
BSMS, MSc, PhD

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

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

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

Data analysis



icmr **NIE**
INDIAN COUNCIL OF
MEDICAL RESEARCH NATIONAL INSTITUTE OF
EPIDEMIOLOGY

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

Data analysis



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

nie.gov.in

HEALTH RESEARCH FUNDAMENTALS

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*



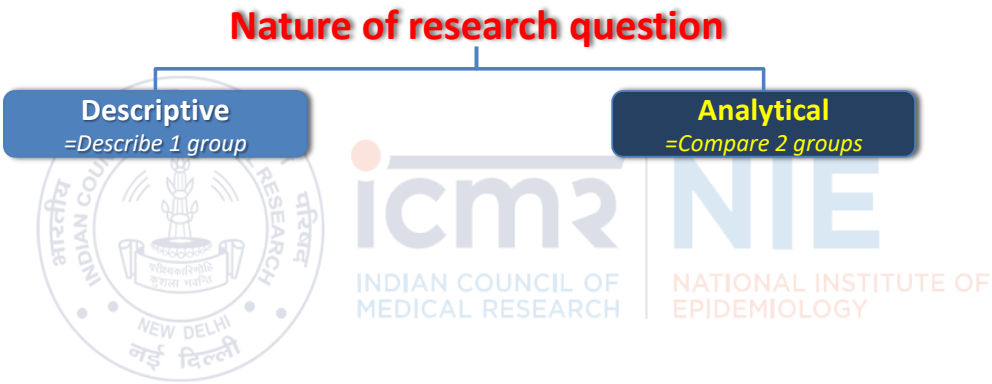
icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

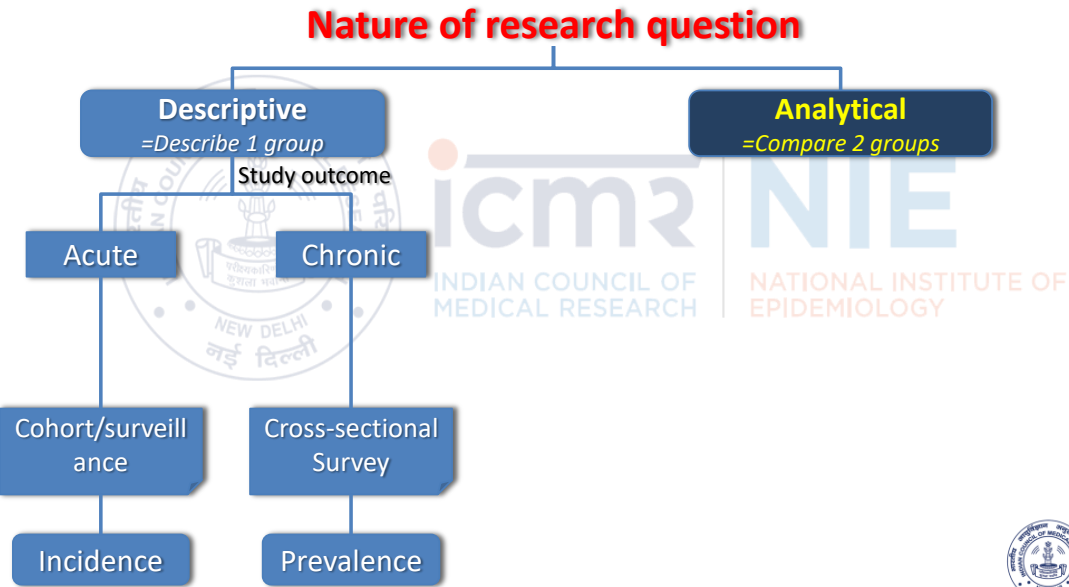
nie.gov.in

HEALTH RESEARCH FUNDAMENTALS

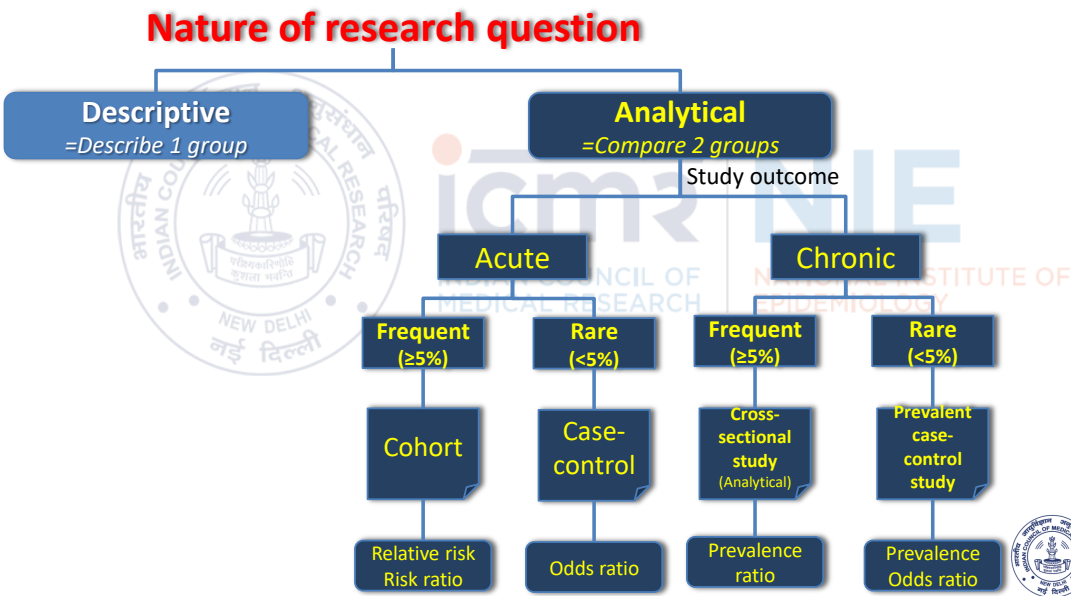
Analysis plan depends on objectives and study type



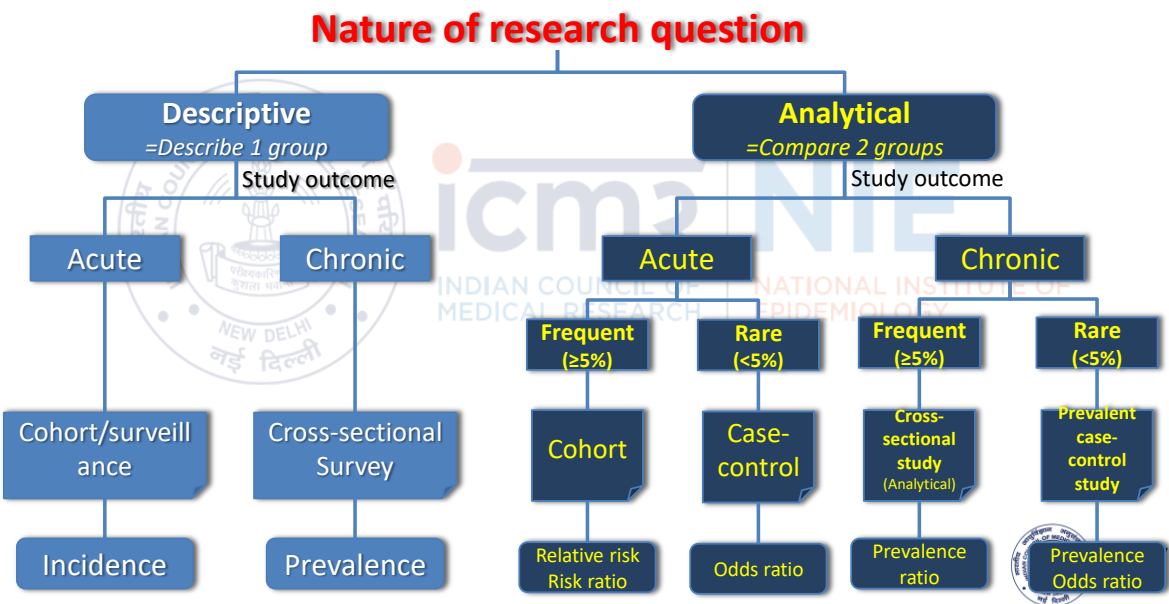
Analysis plan depends on objectives and study type



Analysis plan depends on objectives and study type



Analysis plan depends on objectives and study type



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

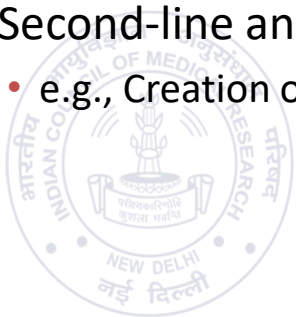
- Baseline characteristics
 - Distribution of study participants by socio-demographic- economic variables
 - e.g., Age, gender, income
- Frequency of clinical features/ health problems
- In analytical study ☐ for compared groups

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



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

7. Conduct advanced analysis

- Dose-response
- Stratifications
- Multivariate analysis



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

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



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCHNIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

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*



*EpiInfo 3.5.4 or EpiInfo 7.1.5; www.cdc.gov/epiinfo

icmr
INDIAN COUNCIL OF
MEDICAL RESEARCHNIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY



icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

icmr
INDIAN COUNCIL OF
MEDICAL RESEARCH

NIE
NATIONAL INSTITUTE OF
EPIDEMIOLOGY

Thank you
