

Sequence of data analysis strategy

1. Identify study type

- Estimation or testing of hypothesis; review the study objectives

2. Identify main variables

- Outcomes:
- Exposures:
- Potential confounders (a priori, before the analysis) (e.g., age, gender, socio-economic status....)
- Variables for sub-group analysis:

3. Become familiar with the data

- Frequency distribution: Examine the frequency of all the variables
- Descriptive statistics: All the variables describing the study population (See point 4 below).

4. Characterize study population

- Demographic characteristics: Distribution of study participants (in case of an analytical study: for the two groups) by age, gender, ...
- Clinical features: If applicable

5. Examine outcome / exposure association on the basis of:

- Hypotheses: In case of an analytical study, compare the two groups for the frequency of exposures (See #2 above) using appropriate measure of association. Variables for which information was collected with multiple levels of exposure will be dichotomized a priori (before the analysis) on the basis of frequency distribution of exposure (e.g., median level of exposure).
- Prior knowledge:
- Study design:

6. Create additional two way tables (On the basis of analysis results)

- New variables: Spell out how you will create new variables (e.g., socio-economic score on the basis of few variables)
- Decide whether you will require additional tables (e.g., results on the basis of weighted analysis; Results adjusted for age/gender)

7. Conduct advanced analysis

- **Dose-response** (for variables associated with outcome in the univariate analysis, whether significant or not): Examine dose-response relationship by using the chi-square for trend (that can be examined with equal categories of exposures) or by using the regular chi-square (for variables for which the data was not available in equal categories of exposures).
- **Stratified analysis:** Conduct stratified analysis to diagnose confounders or effect modifiers (i.e., that were associated with outcome in the univariate analysis)
- **Multivariate analysis:** Determine the type of multivariate analysis required; Determine the variables to be included in the analysis [Outcome (independent) and Exposure (dependent or risk/protective factors); interaction terms]; conduct the multivariate analysis.