

CHAPTER 3:

DATA MANAGEMENT AND ANALYSIS

3.1 Questionnaire, software development & handheld devices

3.2 Data management

3.3 Weighting of data

3.4 Statistical analysis

3.1 QUESTIONNAIRE, SOFTWARE DEVELOPMENT & HANDHELD DEVICES

The questionnaire programming was done through Open Data Kit (ODK), which is an open source application used for development of NNMS application in English and 11 different local languages. It was built-in with quality assurance checks for range and consistency errors. The application used barcode scanning to uniquely identify each household thus, eliminating the typographical errors across various forms. The application also had an option to save partially filled forms as a draft copy. The data was synced by team leaders with centralised online server from all the devices.

3.2 DATA MANAGEMENT

Data cleaning was performed using the *IBM SPSS for Windows version 22.0*. This step included checking for duplicates, missing, or conflicting data and implausible or illogical responses; performing descriptive statistics, frequency tables to look for outliers; checking variables and value labels, for example typing mistakes; ensuring that variables were of the appropriate type (numerical or string); and proper coding for missing data, skip pattern or unanswered questions was followed.

3.3 WEIGHTING OF DATA

Weighting the data was a critical step to have representative results of the entire population of the country. The data from all PSUs were compiled and weighting procedures were followed for adjusting for sampling, population proportions and response rates. For all the steps, data collected from the sample was weighted to provide prevalence estimates at the level of the population, households, area of residence (urban and rural), individuals for age group and gender (adults), only gender for adolescents and urinary samples. (*Annexure 03*)

The detailed statistical analysis plan was prepared based on the identified indicators and subgroups.

3.4 STATISTICAL ANALYSIS

The data analysis was done using *STATA 14.1* with prior developed analysis commands by complex survey analysis. The results of the survey have been presented by descriptive statistics with means and proportions with 95% confidence intervals (CIs) as a measure of precision on the estimated population parameters.