Data Entry & Management

January 28 - 30, 2013

Paramaribo, Suriname

Workshop for Coordinators and Technical Support Persons

Pernell Clarke
Research Specialist
Inter-American Observatory on Drugs
CICAD

Organization of American States

Organization of American States

The Survey Instrument

- The SIDUC survey uses a standard questionnaire
- The majority of questions are coded. The questionnaire contains 95 questions and some have sub-sections with multiple responses
- The questionnaires from which data will be entered must have been reviewed by supervisors.
- The job of the data entry person is to accurately transcribe the information on the questionnaires to MS Excel.

Software: Microsoft Excel

- Microsoft Excel is a computer program that is used to create and manage spreadsheets.
- Within spreadsheets data can be entered managed and organized.
- It allows calculations to be performed.
- It can be used to analyze, and present information in charts, tables, and graphs.
- For the purposes of this project, MS Excel will be used only as a data entry and verification tool.

Pre-entry Data Screening

- Perform all additional coding of completed questionnaires
- Examine questionnaires for response patterns that suggest poor data quality
  Examples
    - the questionnaire not being taken seriously by the respondent (i.e., lack of responses, childish comments written on questionnaire)
    - a majority of questions not being answered (e.g. over half the questions are unanswered)
    - "in-line" responding (in which the respondent appears to be giving the same answer to all questions) or other patterned responses (e.g. "zig-zag" responses)
    - Lack of a valid response for sex and age.
- Scan questionnaires for extreme responses (e.g., reporting use of all drugs frequently).

Verification of Data for Accuracy

- Best to verify 100% of all entered questionnaires by re-entering the questionnaire twice, preferably by a different person
- If costs are limited, verify at least a random sample of 10% to 20% of entered questionnaires to reassure yourself that the error rate is acceptably low
- Data entry methods to assist in attaining accuracy used in Excel
  - Minimum/maximum values allowed
  - Valid values
  - Skip and fill rules

When to Delete a Case

- Completed or Not Completed
  - What is the percentage of all variables that is MISSING or should have been answered but was not?
    - Percentage threshold to take into consideration– e.g., 30%, 50%, 70%
- Sincere or Silly
  - Consistency checks
  - Answer pattern for multiple questions
    - Answering all the same way
  - Answering "Yes" to test question – e.g., "Abyss"

Decision Criteria for Final Sample (TBD)

(1) Did not report the use of a fictitious drug

(2) Completed majority of questionnaire (80%??)

(3) No evidence of exaggerated reporting (e.g., reporting frequent use of all drugs)

(4) Did not have multiple questions answered inconsistently and

(5) Provided valid data for key variables (e.g., sex and age)

Duties of Data Entry Personnel

- To be thoroughly acquainted with the spreadsheet used for data entry

- To take several practice runs in the presence of trainers and/or the Coordinator before assuming duties

- To review questionnaires to ensure that they have been properly corrected and coded

- To bring inconsistencies or errors to the attention of the Coordinator or supervisors

The Data Entry Spreadsheet

- The first 4 rows are used for headings and titles for the spreadsheet, question numbers, and a summary of the actual question
- Questions are arranged in columns
- Questionnaires are arranged in rows
- There are two main areas in the spreadsheet
  - The main area is called the 'worksheet'
  - Below the worksheet is the 'verification area' and this duplicates the worksheet area.

Thank You

Inter-American Observatory on Drugs
CICAD
pclarke@oas.org
http://www.cicad.oas.org