Data Management and Sample Size Calculations

Steps Involved in Data Management

- Create data entry programs
  - Preferably in Epi Info
- Test programs
- Identify and train data management staff
  - One data management coordinator
  - Two data entry staff
- Enter and store data
  - Enter data throughout data collection to catch errors in time.

Data Management Issues

- Data entry screens should be set up in advance of data collection
- Double data entry - ideally two computers needed
- Cleaning begins in the field
- Quality control at all levels (interviewer, supervisor, coordinator)
- Check accuracy of variable coding
- Assess plausibility of results
- Avoid open-ended questions

Epi Info (I)

- Available at no cost at **http://www.cdc.gov/epiinfo/**

- Capabilities:
  - Create questionnaires or
  - Customize data entry process
  - Data entry
  - Data analysis

Epi Info (II)

- Can create epidemiologic statistics, graphs, tables and maps using simple commands:
  - READ
  - FREQ
  - LIST
  - TABLES
  - GRAPH
  - MAP
- Some key features:
  - Data Compare (for double data entry comparison)
  - Teaching exercises
  - Allows analysis and import of other file types

Sample Size Calculations

- Sample size needed depends on:
  - Estimated prevalence of women with the outcome
  - Acceptable margin of error

- Needed:
  - Point estimates
  - Level of accuracy desired: 10% common in high transmission areas; in low transmission areas <=5%
  - If more than one site: design factor, usually put at 2
  - Confidence level: how confident you can be that the answer you got falls within these margins: 95% commonly used

Example: ANC sample size, MALI

- Estimated prevalence of
  - Peripheral parasitemia: 59% (Kassoum Kayentao, personal communication)
  - Hb<11g/dl in all pregnancies: 79%
- Estimated prevalence within 10%
  - Peripheral parasitemia: 49-69%
  - Hb<11g/dl: 69-89%
- Statcalc → Sample size & Power
- Population Survey
- Design Effect=2 (correction for non-random sample)
  - Peripheral parasitemia: 93*2=186
  - Hb<11g/dl: 64*2=128
- Pick larger of two (i.e. 186 pregnancies)