Population surveys

Dr Lale Say
Department of Reproductive Health and Research, WHO

Training Course in Sexual and Reproductive Health Research
Geneva, February 2009

Type of epidemiological studies

- Observational
  - Case-control
  - Cohort
  - Cross-sectional – Surveys
- Experimental
  - Clinical trials
  - Community or cluster randomized trials

Why do a population survey?

- Provide current and detailed socio-demographic, health and other data for households and individuals when such information is not available through routine sources
  - E.g. DHS surveys to collect information on health service coverage where a considerable size of the population do not use health service
- Constructing and monitoring national and regional estimates of characteristics or indicators of interest
  - To collect information representative of population of interest

Basic survey designs

- Cross-sectional survey
  - Data are collected at one point in time
  - From a sample selected to represent a larger population
- Longitudinal surveys
  - Data collected in different points in time from the same sample population, or different samples from the same larger population

Data collection modalities

- Individual interviews
  - E.g., DHS surveys, face-to-face administration of questionnaire
- Other
  - Telephone, web, postal
- Combination
- Selection of a method depends on:
  - Characteristics of population, content of questionnaires, expected response rate, costs

Overview of cross-sectional surveys

- Survey design
  - Research question
- Sample design
  - Sample size, Non-sampling errors
- Survey questionnaire
- Survey cost
- Implementation
- Analysis of survey data

Indicators

- Health
  - *Mortality*
  - *Health*
- Responsiveness
- Financing
- Health System Functions
  - *Coverage*
- Composite Goals

POLICY QUESTIONS

RESEARCH QUESTIONS

WHR

Statistical Annexes

Country Reports

Short Report
Detailed Report
Policy Report

Instrument design

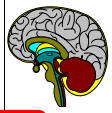
- ✓ Measurement Properties
- ✓ Scales
- ✓ Reliability
- ✓ Cultural comparability

Quality Assurance

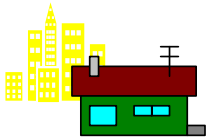
World Health Survey

Statistics

- ✓ Descriptive
- ✓ Multivariate
- ✓ Hypothesis Testing



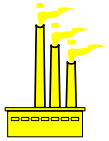
Quality Assurance



Implementation

- ✓ Sampling
- ✓ Training
- ✓ Fieldwork
- ✓ Site Visits

Quality Assurance



Data

- ✓ Editing & entry
- ✓ Checks
- ✓ Cleaning & Filing
- ✓ Missing Data
- ✓ Archiving

Quality Assurance

Survey design – research question

- Set out in advance the aims of the investigation
  - Crucial in defining the "sampling unit" – i.e., an individual or household?
  - If the research question is about one specific individual (e.g., children under 5, then sampling is based on that individual)
  - Most surveys have multiple aims
    - Therefore, households are identified as the primary sampling units

Definition of household – HH

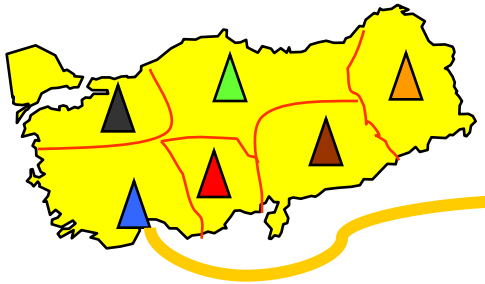
- A household is defined as a single person or group of people who have the address as their MAIN or only address and who either:
  - Share one meal a day (this does not necessarily imply they must always sit down to eat together but that the food is bought or provided on a common basis)

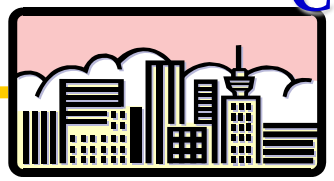
  OR
  - Share living accommodation
- Basically, **they live together and share resources as a unit.**
- Other explanatory phrases can be “eating from the same pot” and “cook and eat together”.

Sampling

- Usually multi-stage stratified sampling design is used
  - Using clusters reduces cost
  - We want to obtain clusters
    - as different from each other as possible
    - including people as similar as possible
  - Stratification
    - Stratifying Factors
      - related to outcome
      - epidemiological variables
      - more often, geographic/administrative variables
- Primary Sampling Unit – units selected for 1st stage of the sampling
  - E.g. cities, census enumeration areas

WHS Sampling-Multistage Clusters – three stage





Counties - PSU s

Provinces-Strata

Enumeration Areas - SSUs

Households - TSUs

Respondent

Factors influencing decision about sample size

- Precision of surveys estimates
- Data quality
- Cost and timeline
- Cluster Sampling
  - increases sample size for fixed budget
  - increases variance
- Design Effect
  - directly influences “effective sample size”

Secondary sampling units (HH selection)

- Household Selection
  - probability inversely proportional to size
  - systematic selection: fixed interval (e.g., every 10th HH in the list) rather than fixed number
  - Need enumeration (listing) of households in sampling unit
    - population registries
    - voters lists
    - other
    - manual enumeration
- Respondent Selection – if there is more than eligible
  - Kish Tables
  - Completion of HH roster
  - Proper documentation of non-response

Data collection instrument - questionnaire

- Instruments include anything from weight scale, to questionnaire, to laboratory instrument
- Questionnaires and other data collection tools must be validated and pilot tested to make sure getting information the you desire
- It is important to adequately define study variables, need a:
  - Good definition
  - Method of measuring it

Questionnaire – variables

- Don’t reinvent the wheel – use/modify established tools whenever possible
  - E.g.
    - Demographic and Health Surveys,
    - World Bank Living Standards Measurement and Social Capital Surveys, and the
    - Centre for Diseases Control (CDC) Reproductive Health Survey
- Modifications to the existing questionnaires usually needed to make them relevant to context

Questionnaire – types of questions

- Open Ended – code later in quantitative research
  - Why did you choose this clinic?
- Close Ended – pre-defined or quantitative numeric answer
  - What is your age?
  - Did you get a Road to Health Card?
- Use close ended as much as possible

Common problems with questionnaires

- Poor questions – unclear, badly worded, complex
- Leading questions
- Handling of sensitive topics (do at end)
- Lengthy recall periods
- Interviewers free to interpret answers
- Too many questions, too long, poorly organized, confusing skips
  - Max 30-35 min.
- Inappropriate or non-standard translations
- No clear assurance of confidentiality

Implementation – Before you go to the field

- Financing & Budget
- Work plan
  - Timeline
  - Filed work logistics
  - Data entry logistics
- Develop instruments
- Drawing a sample of household
- Training manuals
- Pilot test
- Publicity campaign

Conducting the interviews

- Interviewer's name and organisation
- Explanation of research purpose (information sheet)
- Sponsors
- Explanation of how household was selected
- Request to interview eligible person(s) living in the household
  - Informed consent – prior to start
- If sensitive questions – request a private place for asking the questions
- Use first language of the respondent

Validation of data collected

- Review of data collection forms
  - Daily review by study team – provides opportunity to go back the next day for missing fields, etc.
  - Regular intervals by supervisor
  - Prior to data entry
- Tracking of data
  - Logs on site for tracking data to be collected, completed data collection, submission for data entry, etc.
- ALL forms & logs should have date and code, initials or signature of person completing.
- Want processing at each step as soon as possible.

Data cleaning – before data entry

- Query at site – missing data, etc. correct before submit for data entry
- By supervisor – queries, corrections and on-going monitoring of the study team
- ALL corrections made on original (kept at site) with signature & date, original is then copied again and re-submitted to data entry.

Non-sampling errors

- Non-coverage
- Non-response
- Measurement
  - Questionnaire
  - Data collection mode
  - Interviewer
  - Respondent

Survey cost

- Availability of survey infrastructure
- Preparatory activities
  - Training
  - Questionnaires & manuals
- Data collection and filed work
  - Personnel
  - Transport
  - Equipments & consumables
- Results disseminations
  - Printing
  - Seminars

Analysis of survey data

- Define analysis population
- Estimates of indicators – Usually rates
  - Numerators – e.g. N of women aged 15-49 yrs in the sample who use a contraceptive method
  - Denominators – e.g. N of women aged 15-49 yrs in the sample
  - Indicator – contraceptive prevalence rate
- Standard errors of these estimates – to calculate confidence intervals
- Details (overall and by selected characteristics)
- Adjustments
  - weights – for design effect
  - design-based analysis – clustering, individuals within same clusters tend to be similar in explored variables

Thank you !

- More information

  http://www.measuredhs.com

  http://unstats.un.org/unsd/HHsurveys

sayl@who.int