Study Design and Sampling Working Group
Sarah Franklin, Statistics Canada

Second Meeting of the GGP Informal Working Group, Prague, 24-26 February 2003
Overview of Presentation

- Issues with designing a longitudinal survey
- Steps of a survey
  - explain key survey concepts
- Possible guidelines for GGS sample design
Issues with Longitudinal Surveys

- More complex estimation, analysis, organisation
- Requires budget for the life of the panel
- May increase response error
- Can lead to high nonresponse over time
  - Good tracing procedures necessary
- Hard to maintain cross-sectional representativity
Designing a Longitudinal Survey

- Define objectives
- Design questionnaire
- Select survey frame, sample design
- Data collection
- Data capture, coding, editing
- Estimation (weights), analysis
- Dissemination, documentation
Longitudinal vs Cross-sectional Population

- Longitudinal population (one)
  specific population established at a given time point and followed through time (i.e., wave 1).

- Cross-sectional population (many)
  specific population related to a given time point (i.e., wave 1, wave 2, wave 3)
Target vs Survey Population

- **Target population for GGS**
  - non-institutionalized national population aged 18-79.

- **Survey population for GGS**
  - may exclude members of the target population
  - e.g., Residents of Canada on Jan 1, 2004, 18-79, excluding those living in institutions, on reserves, military bases, or no fixed address.
Survey Frame

- Types of survey frames:
  - list (e.g., population register)
  - area

- Frame defects:
  - undercoverage
  - overcoverage
  - duplication
  - missclassification
Sample Designs

- **Probability sampling:**
  - random selection of units
  - probability of selection >0 for all units
- **Examples:**
  - Simple Random Sampling, Cluster Sampling, Multi-Stage Sampling, Stratified Random Sampling.
Simple Random Sampling

- Each population unit has an equal chance of being selected
- Each combination of population units has an equal chance of being selected
Simple Random Sampling

Example:
- \( N = 36, \ n = 12 \)
- Probability of Selection \( = \frac{12}{36} = \frac{1}{3} \)
- Sample Weight \( = 3 \)
Simple Random Sampling
One-Stage Cluster Sampling
Two-Stage Cluster Sampling
Divide population into homogeneous, mutually exclusive groups (strata)

Three main reasons for stratification:
- subpopulation estimates required
- more efficient than SRS
- protects against drawing a ‘bad’ sample
Stratified Sampling
Initial Sample Size Considerations

- Sample size usually determined by required precision (sampling error) of estimates
- Factors affecting precision:
  - variability in the population
  - sample design and estimator
  - survey response rate
  - size of the population
Initial Sample Size by Stratum

\[ n(h) = \frac{n_{\text{respondents}(h)}}{rr_{\text{wave}1(h)} \times rr_{\text{wave}2(h)} \times rr_{\text{wave}3(h)}} \]

\[ n(h) = \frac{500}{.70 \times .70 \times .70} \]

\[ n(h) = 1,458 \]
Questionnaire Design and Data Collection

- Questionnaire design
- Data collection is greatest single expense:
  - listing, tracing, conducting interviews, monitoring quality
- Modes of collection:
  - self-enumeration
  - interviewer-assisted (face-to-face, telephone)
- Computer-assisted applications
Estimation

- Calculate design weights, adjust weights for nonresponse, post-stratification
- Each survey estimate should be accompanied by an estimate of its sampling error (variance)
Data Dissemination

- Analysis of longitudinal data requires access to microdata
- Some countries have specific confidentiality requirements
- International analysis and presentation of GGS results
Possible Guidelines for GGS Survey Design

- Target population: non-institutionalized national population aged 18-79
- Survey population: may exclude up to x% of target population
Possible Guidelines for GGS Survey Design

- Survey frame should:
  - cover target population as much as possible
  - include auxiliary data for sampling (e.g., variables for stratification -- sex, age, subnational regions) and nonresponse evaluation (i.e., demographic variables -- level of education, place of residence, size of household).
Possible Guidelines for GGS Survey Design

- **Sample Design:**
  - 10,000 respondents at wave 3
  - 3 waves, 3 years apart
  - use probability sample design
  - stratification recommended (age, region)
  - allocate sample across strata
  - use design most appropriate to country

- Committee to review all proposed sample designs
Possible Guidelines for GGS Survey Design

- Nonresponse
  - follow-up nonrespondents -- at least 3 attempts
  - adjust sample size for nonresponse (by stratum)
  - agreed-upon tracing procedures
    - capture tracing information on questionnaire
  - include nonrespondents on final micro file
    - all sampled persons must appear on final file
  - perform nonresponse weight adjustment
Possible Guidelines for GGS Survey Design

- Weighting (Longitudinal and/or Cross-sectional), at each wave:
  - calculate sample design weight
  - perform nonresponse weight adjustment at each wave
  - recommend post-stratify to known population totals (e.g., demographic estimates by age, sex)
Final micro file should contain, for each record:

- final status codes (e.g., respondent, nonrespondent, out-of-scope)
- basic demographic information
- design weight
- estimation weight
- replicate weights for variance estimation
Possible Guidelines for GGS Survey Design

- Certification of survey results
  - compare survey estimates with external sources to verify results
Possible Guidelines for GGS Survey Design

- Documentation - sample design report:
  - population covered by survey
  - sampling frame used
  - stratification variables
  - sample size and allocation
  - sample design
  - post-stratification

- Other documentation (e.g., data collection)
GGS Issues

- Sample size: fix precision of estimates?
- Will GGS require cross-sectional estimates?
- Use of area frames by some countries
- Use of telephones for collection
- Weighting and reducing nonresponse bias
- Standardisation of definitions, processing and data collection procedures
- Analysis -- ultimate uses of data
Contact Information

Sarah Franklin
Statistics Canada
Main Building 2500-J
Ottawa, Ontario
Canada K1A 0T6
tel: 613-951-0089
fax: 613-951-0562
email: fransar@statcan.ca