Issues in Survey Research Design

Defining target population, sampling frame and sampling approach
Not as linear or as sequential as this figure suggests...
Why Sample?

• Advantages
  – Cost
  – Time
  – Quantitative
  – Contemporary
  – Replicable

• Disadvantages
  – Superficial
  – Obtrusive
  – Structured-not flexible
  – Based on self report
  – Expensive
Main Factors Affecting Sampling Design Decisions

- Cost constraints
- Time constraints
- Minimizing error
Steps in Sampling

Population of Inference

Target Population (finite group of units to which inference will be made)

Sampling Frame (list/procedure to identify all elements of target population)

Sample (units selected for study with known probability)

Respondents
Population of Inference

Target Population

Coverage Error

Sampling Frame

Sampling Error

Sample

Non Response Error

Respondents
Consider Problems with Sampling Frame

- **Availability**: Is there a frame available or do you need to construct it?
- **Coverage**: Is there a discrepancy between statistics calculated on the sampling frame and the same statistics calculated on the target population
  - Reasons: not up to date, not complete, etc
Components of Coverage Error: Consider the Frame

- **Under coverage**-
  - Omission of units of target population in frame population-Most common form of coverage error

- **Over coverage**-
  - Inclusion of multiple units of frame population attached to one unit in target population i.e. duplicate records

- **Extraneous**
  - Inclusion of foreign units in frame population
When do household surveys undercover populations? Who is Missing?

• The military
• People in institutions—prisons, nursing homes, etc
• College dormitories
• People who may not be listed as household members because they are “residentially unstable,” marginally connected to hh or other reasons (concealment)
• When should we be concerned?
What can be done?
Undercoverage/Overcoverage

• Ignore or disregard problem
• Redefine the target population to fit frame population
• Correct the entire frame
• Use Multiple frames
• Use statistical approaches to re-adjust
Probability Sampling: Sampling Error

Every element in the population has a known non-zero probability of being selected.

Simple random sample SRS–assigns equal probability of selection to each frame element.

Sampling Error-

- Unavoidable variance around point estimates from particular sample – Sampling Variance-reduced with big samples.
Modifications of Simple Random Sample

- Systematic sample
- Stratified sample
- Multistage sample
- Cluster sample
Systematic Sampling

- Sample of size $n$ from population of size $N$ in which the sample units are obtained by a selection interval applied to list
- Random start with every $k$th selected in both directions
- Variance of estimate is biased and not straightforward
Stratified Sample

- Strata are mutually exclusive groups of elements in a frame (15-19 & 20-24 yr olds)
- Separate samples drawn from each group
  - Sampling can be done to accomplish a self weighting sample (based on proportion of elements in population)
  - Sampling can be done to adjust size of subsamples but requires adjust for population estimate (based on desired subsample size-but probabilities of selection still known)
- Reduces variance/increases precision
Cluster Sampling

- A sample in which the sampling units are clusters of samples
- Equal sized or unequal sized clusters
- Often carried out for convenience rather than more precision
- Household surveys usually multi-stage cluster designs
Geographic Area Household Surveys

- Involves
  - Selection of Geographic Units (eg Counties and smaller units within)
  - Enumeration of households within smallest units and selection of households
  - Clustering of households
  - Deduction from household to individual
Design Effect (Deff)

- The ratio of the actual variance of a statistic obtained in a modified sampling design (e.g., clustered sample) to that for the estimator from simple random sample (srs) with the same number of n elements.
- Aiming for close to 1 (srs design effect)
Implications of the Design Effect

For stratified sample deff often < 1 (more precision than srs)

For cluster samples deff > 1 (less precision than srs)

Deff depends on the intraclass correlation coefficient within clusters
  - Complete homogeneity within clusters deff > 1
  - Extreme heterogeneity deff < 1
  - Random distribution within clusters deff = 1

Implications for power analyses to determine sample size: needs to adjust for deff
Examples

• Small Community Based Study
• NYC Hanes
• National Longitudinal Study of Adolescent Health
• National Health Interview Survey + SSA supplement
• Vaccine Coverage study in Ethiopia
• Sexual Behavior in the U.S.
• Head Start Evaluation
Error Associated with Sample Strategy

- Population of Inference
  - Target Population
  - Sampling Frame
    - Coverage Error
    - Sampling Error
      - Sample
        - Non Response Error
          - Respondents