The Multi-Level, Multi-Source (ML-MS) Approach to Improving Data Quality in Surveys

Tom W. Smith, NORC/University of Chicago

ML-MS Approach

- Extract all case-level & aggregate data from sample frame
- Link all sample cases to auxiliary data (AD) sources
  - Case-level
  - Aggregate-level
- Iterative searches of AD
- Paradata

Sampling Frames

- Available data highly variable across sampling frames and these vary across countries
- Population registers; Address-based samples; Random walk
- Case-level
- Aggregate-levels (multi-levels)

Auxiliary Data (AD) - I

- Case-level links by address/other identifiers
    - Found/not found
    - If found, add additional information
- Aggregate-level links
    - Specific address (city-style)
    - GPS readings
    - Postal code
    - Community and other geocodes

Auxiliary Data (AD) - II

- Types of AD

  - Household/individual-based, general databases (e.g. Accurint, Equifax, Peoplefinders)

  - Administrative records – single purpose databases (e.g. voting records, government benefit lists, association membership rolls)

Auxiliary Data (AD) - I

- Increased computerization has made more AD available and eased linkages to surveys

Iterative Searches

- Information gained from initial searches can facilitate additional searches

- Using address-based reversed directory can find out both phone number and/or resident's name

- Name/phone number can then be used in searches of additional databases to make new links and add new information

Paradata

- Process
  - Interviewer characteristics (All)
  - Record of attempts (A)
  - Interactions with household members (A)
  - Keystroke (Respondents)
  - CARI (R)
- Observational
  - Neighborhoods (A)
  - Residence, Exterior (A)
  - Residence, Interior (R)

ML-MS Methodological Uses

- Detect and adjust for non-response bias
- Fill in item non-response for some variables
- Validate information supplied by respondents
  - AD sometimes superior sources
  - AD other times alternative source for cross-checks
- Validation interviews

ML-MS Substantive Uses - I

- Multi-level contextual effects
- Individuals nested within households, neighborhood, communities, regions, nations
- Examples of contextual, aggregate-level geographic effects from General Social Survey

 - racial composition of racial prejudice

 - crime rates on law enforcement attitudes

 - poverty levels and support for welfare

ML-MS Substantive Uses -II

- Making theory-driven, contextual, aggregate-level, geographic data a <u>routine, automatic</u> part of all surveys, not an special, extra (and costly) add-on

Confidentially

- Deductive disclosure risk increased
- Need to have public file edited to insure that deductive disclosure is not possible
- Use of only publically-accessible information from AD sources
- Confidentiality rules vary across countries and target populations

ML-MS Cross-nationally

- Can be applied cross-nationally
- But approach can not be standardized or rigid
- Sample frames vary, available AD sources vary, and rules governing access vary
- Only paradata suitable to standardization
- ML-MS must adapt to those variations

ML-MS Summary

- Augment survey data with data from 1) sample frame, 2) AD, and 3) paradata.

- Include multi-level measures from case-level and aggregate-levels (e.g. neighborhood, community)

- Use for both methodological and substantive purposes

- Make a routine part of doing regular surveys