Westat

Methodological Issues for Longitudinal Studies

Graham Kalton
grahamkalton@westat.com

Varieties of Longitudinal Studies

A. **Cohort studies of individuals**
   - **NLSY, ECLS-B and K, PATH, HRS (financial units), NHATS**

B. **Household panel surveys**
   - **PSID, Understanding Society**

C. **Rotating panel surveys**
   - **SIPP, MEPS, NCVS, CPS**

- **A or B is needed for most longitudinal analyses; these designs are also used for cross-sectional analyses.**
- **C is primarily focused on cross-sectional analyses, but requiring aggregating data across waves and/or a bounding interview.**

General Design Considerations

- National vs. local population
  - Issues of mobility out of area
- Oversampling of domains based:
  - on fixed characteristics (e.g., race, sex, age)
  - not on transient characteristics (poverty status)
- Design with a recognition that objectives will change over time
- Plan for replenishment at some later time to add a sample of new entrants to the population (e.g., immigrants or persons reaching the study's minimum age requirement)

Following Rules: Which Sample Members not to Field in the Next Wave

- Deaths
- Persons who are in institutions: prisons, hospitals, nursing homes, armed forces?
- Persons who have moved permanently abroad (or outside the survey area)?
- Persons who have moved temporarily abroad (or outside the survey area)?
- All nonrespondents? Adamant refusals? Other refusals? Untraced? All nonrespondents at two successive waves?

Factors to Consider with Following Rules

- How many and what types of nonrespondents at one wave are respondents at the next wave?
- Cohort analysis vs. cross-sectional analysis vs. aggregate analysis
- Types of longitudinal analysis
- Simple attrition weighting if nonrespondents at one wave are not fielded in the next
- Otherwise, either a multitude of weights or mass imputation of missing waves:
  - How well can the missing wave data be predicted?
  - Should the imputation be repeated when data for the following wave are available?

How Great is the Sample Loss?

- **Conventional wisdom has been that retention rates are generally high, with wave-on-wave reinterview rates rising to the high 90%'s after a few waves.**
- **However, cross-sectional (initial wave) response rates have been steadily decreasing.**
- **Are reinterview rates remaining high?**
  - **Schoeni et al. (2013) found that they remain high, even increasing in some studies — HRS, NLSY 1979, PSID, BHPS, GSOEP and HILDA.**
  - **Does this hold for other studies?**
  - **What design features make this happen?**

Strategies to Maintain Response (Schoeni et al.)

- Incentives
- Communication with respondents between waves
- Advance notification letters, with number to call
- Use of informants
- Tracking strategies
- Letters to encourage response
- Interviewer contiguity
- Use of opinion leaders
- Interviewer incentives
- Providing information to respondents
- Using proxy informants when needed

Use of administrative records

- Reduce respondent burden
- Obtain data for sample members prior to entering the study
- Obtain data for sample members after they leave the study
  - Also useful data for nonresponse adjustments
- Need respondent's permission.
- Need to assess quality of the administrative data.

Adaptive Design

- **Use survey data and paradata from prior wave to determine protocol for data collection for current wave.**
  - **To reduce costs**
  - **To improve survey quality such as response rate**
- **How effective is this?**

Sample Replenishment

- To incorporate new entrants to the population
  - PSID new immigrants
  - HRS and PATH, persons aging into the study
- To compensate for sample losses
  - Sampling groups with higher nonresponse adjustments at higher rates
- Use of composite estimation to combine old and new sample members: weighting issues
- Two panels: one starting at baseline and the other starting at time of refreshment
- How long should a longitudinal study run before being replaced?
  - Overlapping panels.

Interval between Waves

- Shorter intervals are better for recall, but cost more
- Sometimes desirable to contact panel members at the time of some event in their lives, e.g., a serious illness. Idea discussed at CNSTAT meeting for the follow-up for the NLTCS.
- Intervals of one year are common
  - PSID and HRS started with one-year intervals but now have two-year intervals.
  - Use of event history calendar to date events
- Can administrative records provide data between waves, or to trigger contact with the respondent?
- Can different intervals be used for different subgroups?

Mode of data collection

- In-person
- Telephone
- Mail
- Web
- Mixed modes within a wave and across waves
- Mode effects
- Dependent or independent interviewing
  - How to implement dependent interviewing

Panel Conditioning or Time-in-Sample Bias

- **Many studies and several reviews of panel conditioning effects**
  - **Caporaso et al. reviewed results of 38 studies at AAPOR this year**
- **Effects depend on subject-matter, interval between waves, time in panel, mode and mode changes.**
- **Possible causes: increased awareness of topic, desire to appear consistent, changing motivation over time.**

Analyses of Longitudinal Surveys

- Cross-sectional analysis by wave
- Measures of net change
- Measures of gross change distorted by
  - Differential measurement effects
  - Panel conditioning effects
  - Random measurement error
    - Reliability studies
  - Changes in interviewers
- Seam effects
- Interval between interviews varies.
- Some forms of longitudinal analysis, e.g., growth curves, can handle missing waves.
- Age-period-cohort effects

Statistical Disclosure Control

- **The rich set of data from longitudinal surveys needs to be made widely available for analysts.**
- **The production of public use files for longitudinal studies that preserve the confidentiality of respondents presents major challenges, particularly when the datasets are released on a wave to wave basis**
- **The addition of administrative data and geographical characteristics exacerbates the challenge**
- **Restricted use files and statistical enclaves may need to be the solutions for many analyses.**

Management Issues

- **Large-scale longitudinal studies are national resources.**
- **They need to be guided by a wide range of experts in the study's subject-matter.**
- **This guidance should be on-going in order to make sure the study adjusts to current needs.**
- **Funders need to recognize the major benefits of a longitudinal study, and publications, will emerge as the study matures.**
- **Value of an innovation panel**
  - **Mode effects, adaptive design, incentives, etc.**

Priority Research Issues

- **Initial response rates and retention rates**
- **The effects of changing modes on longitudinal analyses**
- **Reducing respondent burden**
- **The use of administrative records**