Interviewer Observations in Survey Research: Best Practices and Recommended NSF Funding Priorities

Brady T. West, PhD
Research Assistant Professor
Survey Research Center
University of Michigan-Ann Arbor
Interviewer Observations: Some (Very) Loose Definitions

1. Observations recorded by survey interviewers for all sampled units that describe selected features of the sampled units, including attempts at recruitment

2. Observations recorded by survey interviewers for respondents that describe aspects of the survey interview

In both cases, these can be thought of as paradata, or data that describe the process of collecting survey data.
Examples of Interviewer Observations

• **On the Sampled Area**
  – Evidence of Crime / Social Disorder (L.A. FANS)

• **On the Sampled Household**
  – Presence of Young Children (NSFG)

• **On the (Potential) Survey Respondent**
  – Perceived current sexual activity during a screening interview (NSFG)
  – Opinion of quality of data provided, recorded post-interview (PSID)
Why Interviewer Observations?

- Inexpensive source of (potentially) useful auxiliary information on all sample units in face-to-face surveys (RDD surveys also?)

- Existing literature has shown that interviewer observations can be correlated with both response propensity and key survey variables:
  - Useful for Nonresponse Adjustments
  - Useful for Responsive Survey Design (e.g., case prioritization)
Why Not Interviewer Observations?

• Not all interviewers view them as easy to collect or worthy of their time and effort
  – Survey organizations need to emphasize why the observations are important to record...

• Existing literature has shown that they can be error-prone and frequently missing

• Some types of observations (e.g., post-survey) take large amounts of time for interviewers to record: do the benefits of the observations for the larger survey research outweigh the costs?
Best Practices

• Every observation that interviewers are asked to record should have a purpose:
  – Nonresponse Adjustment?
  – Prediction of Response Propensity?
  – Profiling of Active Cases for Possible Targeting?
  – Assessment of Data Quality?

• Collecting observations with no apparent purpose is a waste of time and money
Best Practices, cont’d

• Observations collected on all sample units should be correlates of key variables and / or response propensity
  – Requires empirical studies of these associations
  – Observations can be very useful for informing response propensity models
  – Best observations should be designed to serve as proxies of key measures in the survey
    • NHIS is currently considering to ask interviewers to look for cigarette butts, indications of aid for the handicapped, etc.
Best Practices, cont’d

• Observations collected on respondents should actually be analyzed
  – Shocking to me how rarely this is done on the large surveys at the Survey Research Center
  – These observations can take a great deal of time and effort on the part of the interviewer: shouldn’t they be used to improve operations?
  – These observations may point to problems with the questionnaire in general or indicate data quality issues
Best Practices, cont’d

• When possible, the **quality of observations should be assessed** using validation data
  – Actual survey responses
  – Administrative records
  – Observations by independent interviewers

• Reduced quality has been shown to impair nonresponse adjustments (few studies)

• Interviewers also vary substantially in terms of observation quality, both in U.S. and Germany
Best Practices, cont’d

• Methods are needed for standardizing the ways in which observations are collected
  – Visual examples of how to make effective observations in interviewing training (ESS)
  – Providing interviewers with known and observable predictors of features being observed (NSFG)
  – Asking interviewers to provide open-ended justifications for why they record a particular value for an observation (NSFG)
Open Questions That Need Funded Research

• What is the validity (accuracy) and reliability of interviewer observations across a variety of different FTF surveys, both nationally and internationally?
  • Validation studies require “gold-standard” data
  • This is the area where most of the research has been conducted so far, but...
  • These findings are generally not published in academic journals (proceedings, tech reports, etc.)

• What are the drivers of observation accuracy in different surveys?
  • Respondent, interviewer, and area features
  • Is it pointless to request observations in certain cases (e.g., apartments in urban settings)?
Open Questions That Need Funded Research, cont’d

• What do interviewer observations add to existing auxiliary variables (e.g., purchased commercial data)?
  • Do the observations explain additional variance in survey outcomes, on top of existing variables?
  • Are interviewers really picking up information about sampled units that we don’t already have?
Open Questions That Need Funded Research, cont’d

• What are the impacts (statistical and operational) of reduced observation quality on adaptive survey design strategies?
  – Do decisions based on the observations (e.g., targeting certain cases) backfire because of reduced quality?
  – Or do the observations really serve to improve efficiency?
Open Questions That Need Funded Research, cont’d

• How do error-prone interviewer observations affect various types of statistical adjustments for survey nonresponse?
  • Only weighting class adjustments have been studied thus far...

• What are effective design strategies for improving observation quality?
  • Provide interviewers with known auxiliary predictors of features being observed to help them, or just use the predictions?
Open Questions That Need Funded Research, cont’d

• How can post-survey observations be used to improve survey estimates?
  • Calibration estimators (Kott, 2006, SM)?
  • Indicators of data quality (i.e., analysts might consider dropping certain cases)?

• What are the sources of interviewer variance in observation quality?
  • Qualitative study of different strategies used to record observations in the field...
  • Actually TALK to the interviewers about what they do!
Open Questions That Need Funded Research, cont’d

• What are the empirical trade-offs between the costs of collecting various observations vs. improvements in survey estimates from collecting the observations?
  • Is this practice really worthwhile?
Important Considerations

• Funding this research will remain important as long as the U.S. government (and other agencies) conduct large FTF data collections

• If FTF mode “dies out” (unlikely...), these issues may no longer be relevant (but the quality of other auxiliary variables, e.g., Google earth observations, may become relevant)

• Some studies have looked at the quality of telephone observations (e.g., guessing gender of possible respondents)
Discussion...

- Questions?
- Thoughts?
- Thank you for inviting me!
- Email: bwest@umich.edu