Responsive Survey Design for Fertility Surveys: Paradata-Guided Active Interventions in the National Survey of Family Growth

Robert M. Groves, Nicole Kirgis, Emilia Peytcheva, James Wagner, William Axinn, University of Michigan, USA; William Mosher, US National Center for Health Statistics

1. Introduction

Demographic research on fertility behavior over the past six decades has been tremendously prolific, advancing empirical information, theories of fertility behavior, public policy and programs aimed at enhancing fertility outcomes. The vast majority of this research is based on survey data. These include some of the largest and most advanced social surveys ever conducted, such as the World Fertility Surveys, the Demographic and Health Surveys, and the Family and Fertility Surveys, as well as dozens of national, local and longitudinal surveys around the world. The longest standing series of fertility surveys are the American fertility surveys featuring the National Survey of Family Growth (NSFG) and its predecessors (Thornton and Axinn 2004). The NSFG has been an important leader in methodological advances for the study of fertility using survey data (Mosher 1982, 1988; Thornton and Axinn 2004). This paper describes and documents the most recent NSFG advances, from NSFG Cycle 7, with continuous field work currently ongoing.

Fertility survey data collection is fraught with uncertainty. Some sample frames used have unknown eligibility (e.g., address frames contain some unknown number of age-eligible households). Interviewers are not generally knowledgeable about the times of the days and days of the week that households are at-home. The likelihood of obtaining a consent to be interviewed cannot be predicted well. Unfortunately, each of these uncertainties implies lack of control over the cost, timeliness, and ultimately the error structure of fertility survey data.

If there were no time or cost constraints of survey data collection, these uncertainties would pose no large problems. However, survey researchers always work under those two constraints. Those constraints and the uncertainties mentioned above combine to challenge wise expenditures of time and costs. One implication is that complete specification of the survey design *prior* to the data collection is unlikely to be optimal (it, perforce, is completed prior to exposure to the uncertainties). If one assumes a beneficial realization of the various uncertainties, overruns of time or costs are the key risk. If one assumes a perverse realization of the various uncertainties, smaller samples of respondent cases are the risk.

Most survey design texts assume that design features are fixed prior to the initiation of data collection. These include sample design, sample size, length of data collection period, number of interviewer hours, travel costs, and, sometimes, number of calls to cases prior to the first contact, number of contacts, type and level of calling following a sample person expressing some reluctance to participate. In the presence of the uncertainties reviewed above, some of these may be candidates for mid-data collection alteration.

The argument above is the basis of the use of so-called "responsive designs," which alter the data collection protocol base on paradata collected during the data collection period (Groves and Heeringa, 2005). By way of definition, responsive survey designs:

a. pre-identify a set of design features potentially affecting costs and errors of survey statistics;
b. identify a set of indicators of the cost and error properties of those features;
c. monitor those indicators in initial phases of data collection;
d. alter the active features of the survey in subsequent phases based on cost/error tradeoff decision rules; and
e. combine data from the separate design phases into a single estimator.

In some sense, "responsive design" notions have been part of good survey data collection practice for decades. However, little of this is documented in the literature, and there is no existing conceptual framework for this work.

A potential valuable tool for responsive design is paradata (Couper, 1998) -- process and administrative data that are produced auxiliary to the survey data collection. With computer-assisted data collections, the software that captures the data can also capture interviewer observations, attributes of call attempts on sample cases, nature of the interaction with the household members, and behaviors during the interview measurement itself (e.g., timing of different behaviors, digital audio recording of voices of interviewers and respondents).

This paper reviews a set of paradata that have been monitored over the past year to assess the performance of a face-to-face survey data collection. It describes the conceptual framework of the paradata, illustrates the monitoring tools, and then evaluates two interventions in the data collection designed to alter the course of the data collection in a fashion judged desirable by the survey managers. The results demonstrate how paradata and responsive survey design can be used to yield more and higher quality fertility survey data within the same cost and time constraints. These tools offer a means to dramatically advance the collection of fertility survey data world wide and have the potential to be applied to survey data collection across a wide range of demographic topics.

2. Survey Design

The National Survey of Family Growth (NSFG) is an area probability sample of households, attempting to interview one randomly-selected person aged 15-44 in each sample household. The data collection is continuous throughout the year, consisting of four quarterly replicates with a 12 week data collection period. The twelve week period is divided into two phases, the second of which (weeks 11-12) is a subsampling of remaining active cases at the end of the first phase. The interviewing has two parts – a screener interview that uses a household roster to identify any age-eligible persons (aged 15-44) and a main interview (lasting 60-80 minutes). CAPI and ACASI are employed.

Many paradata are selected for monitoring throughout the survey period. As shown in Figure 1, the paradata produce daily monitoring indicators that are used to examine the changes in performance of a production function that relies on interviewer effort, applied to a set of active sample cases that produces interview data records, whose property of balance across key socio-demographic domains is a key performance indicator. This paper examines the following: calls made to screener cases and to main cases, contact rates of screener cases, screener interviews produced, response rates for key domains assigned different probabilities of selection, discrete hazard models predicting the likelihood that the next call will be an interview (separately for screener and main cases).

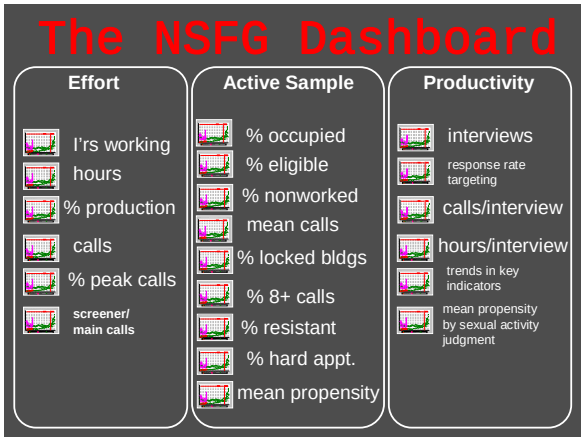

Figure 1. Paradata Indicators Monitored Daily in the NSFG, Cycle 7

3. Methods

Two interventions based on paradata are described and evaluated. The first is typically initiated in the fifth week of the 10 week phase 1 data collection, in which interviewers are asked to focus their efforts on completing screener interviews relative to completing main interviews. We compare the numbers of screener calls and number of screener interviews for the week after the intervention, relative to the two weeks before and two weeks after the intervention. This intervention relies on tracking the ratio of calls made to screener cases to those of main cases, as shown in Figure 2. In weeks 5-6 the intervention yields a jump in the ratio.

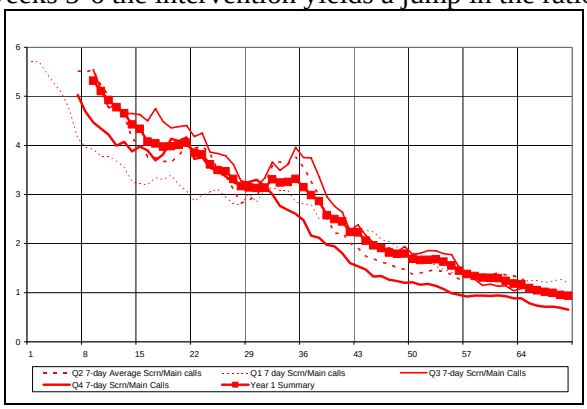

Figure 2. Ratio of Screener Calls to Main Interview Calls by Day by Quarter

The second intervention occurs toward the end of the first 10 weeks of data collection, in which active cases (both screener and main cases) that have large selection weights and have high expected probabilities of an interview on the next call are targeted for greater focus. We compare the number of calls and interview rate for those targeted cases with cases that have similar weights and propensities.

4. Results

Management intervention to increase calling on screener cases increases such volume by 7-25%. Correspondingly, production of screener interviews increases by 10-20%. While the magnitude of the change varies over the four quarters of replicated interventions, the direction of the changes during these interventions is uniform.

When high weight, high propensity cases are targeted, the targeted screener cases received 29% more calls; the targeted main cases, 176% more calls. In the week of the intervention, screener cases in the intervention cell produced a 32% response rate, and those in the adjacent cells, 29%. Main cases in the intervention cell produced a 30% response rate and those in the adjacent cells, 19%.

We address the impact on key survey estimates of the NSFG by examining pre- and post-intervention estimates.

We conclude that the collection of paradata, coupled with active monitoring during data collection and experimental intervention, can yield estimates that have better error properties.

References

Couper. M. 1998. Measuring Survey Quality in a CASIC Environment. Proceedings of the Survey Research Methods Section of the American Statistical Association, 41-49.

Groves, R., and S. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Nonresponse and Costs," *Journal of the Royal Statistical Society, Series A*, 169, pp. 439-457.

Mosher, W. D. 1982. "Fertility and Family Planning in the 1970s: The National Survey of Family Growth". *Family Planning Perspectives* 14(6):314-20.

Mosher, W. D. 1988 "Fertility and Family Planning in the United States: Insights from the National Survey of Family Growth". *Family Planning Perspectives* 20(5):207-17.

Thornton, A and W. Axinn. 2004. "The Social Demography of Fertility: Some Contributions of American Surveys". Pp. 281-304 in *A Telescope on Society*. House, J. S., F. T. Juster, R. L. Kahn, H. Schuman, and E Singer, Eds. Ann Arbor: University of Michigan Press.