Responsive Survey Design, Demographic Data Collection, and Models of Demographic Behavior

ABSTRACT

Falling response rates have led survey methodologists to pioneer innovative ways to use process data ("paradata") to address non-response by altering the survey design. By improving representation of reluctant respondents, responsive survey design also changes our understanding of substantive issues studied by demographers. Using the National Survey of Family Growth Cycle 6 we illustrate how responsive survey design can improve both demographic estimates and models of demographic behaviors based on survey data. By juxtaposing measures from regular and responsive data collection phases, we document characteristics of the general population that are systematically under-represented in surveys not taking special effort to interview reluctant respondents. Using models established in highly cited papers based on NSFG data, we demonstrate how adding reluctant respondents through responsive survey design changes model estimates. Results demonstrate the wide potential of responsive survey design to improve the quality of science in demographic research based on survey data.

Responsive Survey Design, Demographic Data Collection, and Models of Demographic Behavior

EXTENDED ABSTRACT

Introduction

The representative sample survey has been a fundamental foundation of demographic research throughout the history of the discipline. Many of the field's key advances in both empirical evidence and theoretical reasoning are founded on information from surveys. But even as sophistication of both survey measurement and survey analysis dramatically advances, the general population's growing reluctance to participate in surveys across recent decades poses an enormous threat to the field. This problem is greatest in relatively rich countries of Europe and North America, but it is growing across the world. The problem has been documented in detail (Groves and Couper 1998), but demographer's standards for acceptable survey response rates continue to drop, and social scientists devote increasing effort to the study of the consequences of non-response for the substantive issues they investigate. In the midst of growing awareness of a looming scientific crisis, data collection methodologists have pioneered innovative approaches for using newly available data collection technologies to address the non-response problem. Together these approaches are termed "responsive survey design" and they can be used to simultaneously improve survey representation of reluctant respondents and control costs of data collection. Here we describe the application of responsive survey design to a key demographic survey in the United States – the National Survey of Family Growth. Using the NSFG example, we demonstrate how responsive survey design can be used to improve both demographic estimates and models of demographic behaviors based on survey data.

Computerization of the survey data collection process was the essential technological shift that allows responsive design to be used. Though "paper and pencil" data collection continues to be used in some rural parts of the world, "computer-assisted personal interviewing" or CAPI is now used by the majority of the world's demographic survey data collections. Use of computer software for questionnaires promoted data collection instruments that could be more easily tailored to respondents' unique circumstances or previous responses and also allowed for dynamic error detection during field work and more rapid release of data in electronic form. All of these are desirable in the creation of new demographic data. But computerization also provided the means for the creation of survey "paradata", or data about the data collection process itself (Couper 1998). Now, paradata from CAPI data collection, combined with internet technologies that allow paradata to flow from decentralized data collection staff to centralized management, provide the means to centrally manage responsive survey designs in large scale demographic surveys.

The National Survey of Family Growth (NSFG) Cycle 6 (2002-03) featured CAPI interviewing, collection and analysis of paradata, and responsive design on a large scale. The study involved more than 12,500 personal interviews collected nationwide by a staff of more than 300 interviewers. Specifically, the study was designed in multiple phases – a main phase, designed following protocols established before data collection began, and a responsive phase, designed explicitly to use analyses of paradata to direct changes of protocol targeted to improve representation of reluctant respondents (Groves et al 2005). By juxtaposing measures from these two data collection phases, we are able to document characteristics of the general population that are systematically under-represented in surveys that do not take special effort to interview reluctant respondents. Then, using models established in highly cited papers based on NSFG

data, we are able to demonstrate how the addition of reluctant respondents through responsive survey design changes what we learn from demographic models based on survey data. Together this body of evidence shows how responsive survey design provides a new methodological tool to improve the quality of science in demographic research based on survey data. We close by describing the recent evolution of responsive survey design and discussing the wide potential of these techniques to improve demographic data collection.

Responsive Survey Design

Declining response rates present survey designers with increased uncertainty about the performance of their survey design, increased effort required to obtain interviews, and thus increased costs of data collection (de Leeuw and de Heer 2002; Groves and Couper 1998). The development of computer-assisted methods of data collection has provided survey researchers with tools to capture a variety of process data or "paradata" about the data collection process (Hapuarachchi et al 1997; Couper 1998; Scheuren 2001). Paradata can be used to change the design during the course of data collection, in efforts to achieve response rate targets, or lower survey errors and costs. The responsive use of paradata to modify the design during the field period has been labeled "responsive design" (Groves and Heeringa 2006).

Theory and Process of Responsive Design. A key component of responsive design is that it is organized around "design phases" (Groves and Heeringa 2006). The theme of this paper is that the entire survey process, from design through data collection, should be responsive to both anticipated uncertainties that exist before the survey data collection begins and to real time information obtained throughout the survey data collection.

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.

Figure 1 illustrates the key components of a three phase responsive design, in which the first phase is mounted with N design options applied simultaneously (possibly on different replicate subsamples). Examples of these design options might include whether an incentive is offered, the number of follow-up calls to nonrespondent households, the use of a short or long version of a questionnaire, or alternatives for the number of sample persons to select per household. During Phase 1 (as displayed at the bottom of Figure 1) paradata (Couper, 1998) are collected to inform the researcher of the interviewer hours spent calling on sample households, driving to sample areas, conversing with household members, and interviewing sample persons. The paradata may include observations about the characteristics of housing units (e.g., whether they have some access impediments) or utterances of contacted sample persons predictive of later actions, etc. Supplementing the paradata are key statistics from the survey analyzed as functions of interviewer effort, computed on intermediate data sets as interviews are completed.

(*Figure 1, about here*)

At the end of Phase 1, the researcher makes a decision about the Phase 2 design options that appear to be prudent (the middle portion of Figure 1). This decision will be guided by the

paradata information on costs and sensitivity of values and standard errors of key statistics. Phase 3 is often a phase introduced to control the costs of the final stages of data collection while attaining desirable nonresponse error features for key statistics. This might involve a second phase sampling of remaining nonrespondents, the use of different modes of data collection, the use of larger incentives, etc. After the third phase is complete, the survey data collected in all three phases are combined to produce the final survey estimates.

Applying Responsive Design in the NSFG. The field work for Cycle 6 of the NSFG was organized in two distinct phases of operation. The main data collection phase occurred during an 11-month period from March 2002 through January 2003. During these initial phases paradata were collected to monitor information about the data collection. Paradata included items of data such as interviewer performance, observations on neighborhoods and housing units, day and time of call attempts, and observations on contact with household members (e.g. whether they asked a question about the survey or responded with a negative statement about the survey). These paradata were used to build predictive response propensity models – logistic models predicting the odds that the next call on a sample case would produce an interview, given a set of prior experiences with the sample case (for a full description of the paradata collected and their use in propensity models see Groves and Heeringa 2006). Expected propensities were summed over all cases and used to group cases into quartiles, which formed strata for the responsive design phase sample. Thus on the basis of these propensity models, a phase 3 sample with high probability to be interviewed was selected.

The responsive design phase occurred during the last month of fieldwork—February 2003. For this phase the recruitment protocol was altered in attempt to attract sample people on

whom the earlier phase protocols were not effective. The responsive phase recruitment protocol entailed use of the most productive interviewers on staff, increased use of proxy informants for the screening interview to lower the burden of obtaining screener information, a small prepaid token incentive (one-eighth of the main interview incentive as compared to no prepaid incentive in the main phase) for completing the screening interview, and promise of additional incentive (double the main incentive) for completing the main interview. The responsive phase was successful in increasing the overall response rate, by recruiting a large number of respondents who failed to participate in initial phases.

Evaluating Responsive Design in the NSFG. The second phase of data collection, in responsive design, adds cases to the data base, necessarily improving the overall response rate in the study. The American Association for Public Opinion Research has published a standardized set of guidelines for determining the overall response rate of a study that includes a phase of responsive design, weighting those cases to provide a precisely appropriate response rate calculation (to learn more about those calculations for the NSFG, see Groves et al. 2005).

But a key question remains: exactly how different are the cases added to the study through responsive design? Responsive design brings more respondents into the study, but those additional respondents only change what we know if they are different from respondents in the main study in some important ways. Moreover, if they are different in ways that are closely related to key substantive topics being studied adding the cases is more important than if they are different, but those differences are essentially orthogonal to the topic being studied. So evaluation of the substantive consequences of responsive design depends greatly on assessing the substantive differences between respondent to the main study and respondents to the responsive design phase.

One way to assess these differences is to simply compare the characteristics of main study respondent to characteristics of responsive design phase respondents. One might wish to compare their basic demographic characteristics, such as age, race, employment status or education. In general, theories of nonresponse emphasize that busy people are less likely to participate in surveys and are hard to locate (Groves and Couper 1998). This principle yields predictions regarding the expected difference in characteristics for the two phases of data collection – in general the responsive design phase respondents should be characterized by life circumstances that make them busier than main study respondents. So, for example, with regard to employment, we would expect responsive design phase respondents to be somewhat more likely to be employed full time than main study respondents. For a study like NSFG, comparing characteristics closely associated with the substantive aim of the study, such as marital, childbearing or sexual behavior and attitudes, is even more essential. Again, those situations likely to produce busier life circumstances, such as childrearing, are expected to be somewhat more prevalent in the responsive design phase.

Another way to assess these differences is to estimate models of demographic behavior with and without cases from the responsive design phase. This approach is more complex than a simple comparison. It implies a full multivariate model of an important demographic outcome, built as closely as possible to the specifications produced by previous research, with known expectations for values of key parameters. By estimating such a model once with data from the main study and second time with data from the responsive phase, one can capture a heuristic sense of the differences in substantive conclusions likely to result from adding reluctant respondents to the study by using responsive design. Though formal tests of the statistical significance of differences in parameters across these models fit to these two different data

sources remain beyond the scope of currently available analytic tools, inspection of the results provides at least an informative view of likely consequences of implementing responsive design.

In the paragraphs below we evaluate responsive design using both approaches. First we compare the characteristics of respondents for the main study with characteristics of respondents for the responsive phase. Then we estimate multivariate models of key NSFG outcomes twice – once using data from the main study and a second time using data from the responsive phase. The models themselves are based on models published in highly cited work using previous rounds of NSFG data.

Data and Methods

Data for this study were taken from the National Survey of Family Growth (NSFG), Cycle 6. Fieldwork for the NSFG, conducted between March 2002 and February 2003, was done by professional female interviewers who questioned 12,571 men and women ages 15 to 44 in their homes. The NSFG obtained detailed information on factors affecting childbearing, marriage and parenthood.

For these analyses, we focus on two groups among the respondents: those interviewed during the main data collection phase and those interviewed during the responsive design phase. Furthermore, we subdivide these groups by gender, so that our sample contains men interviewed during the main phase (n = 4,601), women interviewed during the main phase (n = 7,146), men interviewed during the responsive phase (n = 327), and women interviewed during the responsive phase (n = 497). Of course the responsive phase of the study is, by design, a small proportion of the total interviews collected. One consequence is limited statistical power for testing differences between the main and responsive phases. This limitation prevents us from detecting small differences and focuses instead on large differences across phases.

Results Evaluating Responsive Design in the NSFG

To evaluate the use of responsive design in the NSFG we examine three different empirical dimensions of the NSFG Cycle 6 data. The first dimension compares basic demographic characteristics of those respondents interviewed in the main study to those interviewed in the responsive design phase. The second dimension compares attitudes and behavior across substantive domains of greatest importance to the NSFG. The third dimension compares estimates from demographic models estimated with and without the respondents added to the study through the responsive design method.

Comparing Demographic Characteristics. In Table 1 we present demographic statistics by interview phase and gender. The numbers presented are percentages, and we test whether the percentages in each category of each variable in the main phase are different from the same percentages in the responsive phase. The overall pattern of age differences is that the responsive phase sample is older than the main phase sample. For both males and females, the responsive phase sample is significantly less likely to be under age 20, and more likely to be age 30 or older, though the difference in this category attains statistical significance only for females. Furthermore, the proportion of Hispanics in each interview phase is significantly different for both genders (21% of females in the responsive phase compared to 14% of females in the main phase, and 25% of males in the responsive phase compared to 16% of males in the main phase). Finally, there's a strong difference in labor force participation but only among females. Females in the responsive phase are much more likely to be employed full time compared to females in the main phase. These measurements provide preliminary evidence that the recruitment strategies adopted during the responsive phase recruited a different type of sample.

(*Table 1, about here*)

Comparing Substantive Attitudes and Behavior. We continue comparing samples from the different interview phases in Table 2. As the results show, the samples display differences in patterns of marriage and childbearing behavior. Strong differences in marital status pertain only to females: Females in the responsive phase are significantly more likely to be married (54% compared to 45% in main phase). Interestingly, responsive phase females are more likely to have ever cohabited than main phase females, but responsive phase males are less likely to have ever cohabited than main phase males. In terms of lifetime number of sexual partners, responsive phase males and females tend to have had more opposite sex sexual partners than main phase males and females. Finally, in terms of childbearing, there is clear evidence that responsive phase males are more likely to have biologically fathered a child than main phase males, and some evidence that responsive phase females have experienced more live births than main phase females.

(*Table 2, about here*)

In Table 3 we present differences in attitudinal measures between the samples. Respondents were read: "It is all right for unmarried 18 year olds to have sexual intercourse if they have strong affection for each other" and given answer choices of "Strongly Agree," "Agree," "Disagree," and "Strongly Disagree." Although the "Neither Agree or Disagree" response was not offered, it was accepted as a response if the respondent insisted. Other attitudinal measures include responses to "Gay or lesbian adults should have the right to adopt children," "It is better for a person to get married than to go through life being single," "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work," and "The rewards of being a parent are worth it, despite the cost and

the work it takes." Responses to these statements were coded in the same way. For parsimony, we compare the samples on one response category for each variable.

(*Table 3, about here*)

Table 3 shows that the samples consistently display differences in responses to attitudinal measures. In response to "It is all right for unmarried 18 year olds to have sexual intercourse if they have strong affection for each other" both responsive phase males and females are more likely to disagree than their main phase counterparts. In response to both statements "Gay or lesbian adults should have the right to adopt children" and "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work" the differences by phase are stronger for men. Men in the responsive phase agree less and disagree more than main phase men. While there are no significant differences between the male samples in response to "It is better for a person to get married than to go through life being single," there are differences among the female samples: responsive phase females are more likely to agree than main phase females. Sample differences in response to "The rewards of being a parent are worth it, despite the cost and the work it takes" are also stronger among females than males. Responsive phase females are less likely to strongly agree than main phase females.

Clearly we have captured a range of differences between respondents recruited during the main fieldwork and respondents recruited during the responsive phase, when higher recruiting efforts were made. Overall we have shown that the sample from the responsive phase is older and higher proportion Hispanic. For females specifically, the sample from the responsive phase is more likely to be employed and married. For males specifically, the sample from the responsive phase has had more sexual partners and is more likely to have fathered a child.

Finally, on a number of attitudinal measures, the responsive phase samples of both males and females seem somewhat more socially conservative.

Comparing of Demographic Models. Our next step will be to estimate models that are common in social demographic literature, with and without the responsive phase sample included in the total analysis sample. We expect that including the responsive phase sample, who would have otherwise been nonrespondents, will change the model estimates because of their differing characteristics from those respondents recruited normally. *To be completed….*

Recent Advances in Responsive Design and Prospects for Demographic Data Collection

To be completed…. Will emphasize new responsive design approaches in NSFG Cycle 7

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 Couper, M. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.

Groves, Robert M. and Heeringa, Steven G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. J.R.Statist.Soc.A, 169(3), 1-19.

Groves, Robert M.; Benson, Grant; Mosher, William; Rosenbaum, Jennifer; Granda, Peter; Axinn, William; Lepkowski, James; Chandra, Anjani. (2005). *Plan and Operation of Cycle 6 of the National Survey of Family Growth*. Hyattsville, MD: National Center for Health Statistics, Vital Health Statistics, 1(42).

Hapuarachchi, P., March, M. and Wronski, A. (1997). Using statistical methods applicable to autocorrelated processes to analyze survey process quality data. In *Survey Measurement and Process Quality* (eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin), ch.26, pp.589-600. New York: Wiley.

de Leeuw, E. and de Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In *Survey Nonresponse* (eds. R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little), ch.3, pp.41-54. New York: Wiley.

Scheuren, F. (2001). Macro and Micro paradata for survey assessment. *United Nations Work Session Statistical Metadata*, Washington DC.

Figure 1. An Illustration of a Three Phase Responsive Design

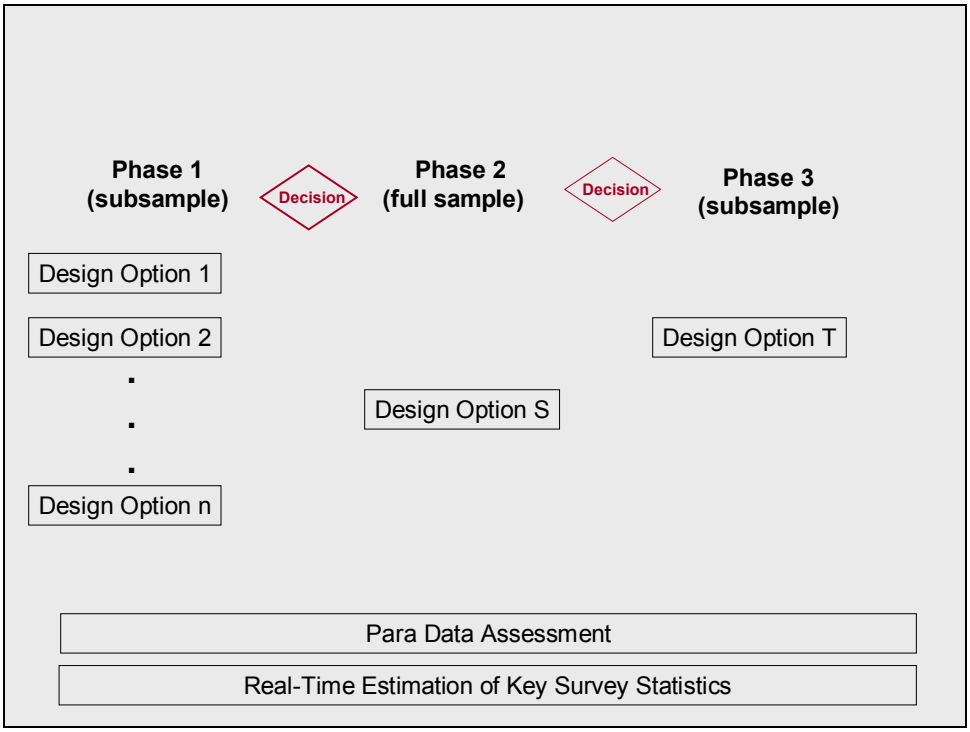

Table 1. Descriptive statistics, NSFG Cycle 6, by interview phase, weighted

	Female		Male	
variable	**Main Study**	**Responsive Phase**	**Main Study**	**Responsive Phase**
Age				
Under age 20	16.55	11.88*	17.46	11.06*
Age 20-29	31.19	29.69	31.02	32.97
Age 30 or older	52.26	58.43*	51.52	55.97
Race				
Black	15.36	13.04	13.53	13.09
White	76.37	77.69	76.23	76.70
Other	8.27	9.27	10.25	10.21
Hispanic				
Yes	13.89	21.25**	15.56	24.74**
Labor force participation				
Full time	42.04	53.80***	60.95	64.20
Education				
More than high school/GED	50.59	52.06	45.53	45.22

+ $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$-Rao-Scott Chi square tests

Table 2. Behavioral Measures, NSFG Cycle 6, by interview phase, weighted

	Female		Male	
variable	**Main Study**	**Responsive Phase**	**Main Study**	**Responsive Phase**
Marital Status				
Married	44.85	54.31***	41.83	44.92+
Ever Cohabited				
Yes	49.68	52.57*	30.14	26.26*
Age at first sex				
Age 20 or older	15.50	20.85*	14.38	18.84+
Number of Male (Female) Sexual Partners in Lifetime				
None	13.68	12.50	13.49	10.76*
7 (7 or more) partners	3.59	5.87***	37.92	42.12*
Ever Biologically Fathered a Child				
Yes	n/a	n/a	45.82	54.54***
Total Number of Live Births				
2 babies	21.38	24.54*	n/a	n/a

+ $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$- t-tests

Table 3. Attitudinal Measures, NSFG Cycle 6, by interview phase, weighted

	Female		Male	
variable	Main Study	Responsive Phase	Main Study	Responsive Phase
Ok if unmarried 18 year olds have sex if strong affection				
Disagree	30.46	37.35***	27.78	33.1**
Gay adults should have the right to adopt				
Agree	41.14	43.88+	39.55	30.19***
Working mother can establish warm relationship with child				
Disagree	14.00	12.24+	23.02	26.54*
Better to get married than go through life being single				
Agree	35.56	40.32**	44.06	45.67
Rewards of being a parent are worth it despite the cost				
Strongly Agree	59.39	55.29**	51.19	54.45+

+ $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$- t-tests