Nonresponse Bias Analyses for the Continuous NSFG, 2011-2015

By
James Wagner, Ph.D. NSFG Chief Mathematical Statistician,
and Mick P. Couper, Ph.D., NSFG Project Director,
University of Michigan

Executive Summary

This brief appendix describes our approach to nonresponse bias analysis and management on the NSFG. Further details on the design and conduct of the NSFG are available in NCHS reports (Groves, et al. 2009; Lepkowski, et al. 2013) and in other publications (Wagner, et al. 2012). The NSFG is conducted using continuous interviewing, with four 12-week quarters per year. For the first 10 weeks of each quarter, we use real-time paradata to manage the survey fieldwork—to direct interviewer effort to where it is needed (e.g., screeners, if not enough screeners are being done; or to Hispanic adult males, if their response rates are lagging). We also use paradata to select cases and structure effort in the last 2 weeks of each 12-week quarter, where we subsample unresolved cases for additional effort. Our goals include equalizing response rates within sub-groups by age, gender, and race, and monitoring the estimates for some key variables from the survey (e.g., percent who have never been married). Paradata are also used to adjust the sampling weights for nonresponse. The overall goal of this design is to manage fieldwork effort on an ongoing basis with the aim of measuring and minimizing nonresponse error for a given level of effort.

The document describes our activities in the 2011-2015 Continuous NFSG. These activities have built upon the success of the 2006-2010 Continuous NSFG, using essentially the same design, but with continuous improvements as more is learned. We continue to improve on the monitoring of daily paradata with a view to further minimizing nonresponse error.

Introduction

As with most large complex surveys in the U.S., the NSFG anticipates a response rate below the 80% target set by OMB. As of this writing, we are in the fourth year of production for the 2011-2015 Continuous NSFG. We have recently released a public-use data file that includes data collected from 2011-2013. In this report, we will review results from 2011-2013 as well as

ongoing efforts to monitor and control nonresponse bias in the NSFG.

Given that NSFG is based on an area probability sample, only limited frame information (other than aggregated census data for blocks or block groups) is available to explore nonresponse bias. Further, given the topic of NSFG (fertility, contraceptive use, sexual activity, and the like), little external data exists to evaluate nonresponse bias for key NSFG estimates. However, managing the data collection effort to minimize nonresponse error and costs is a key element of the NSFG design, and relies on paradata collected during the data collection process to monitor indicators of potential nonresponse bias.

We have the following types of data to assess nonresponse bias in NSFG, which we will discuss in turn, below:

1) a paradata structure that uses lister (usually an interviewer) and interviewer observations of attributes related to response propensity and some key survey variables;
2) data on the sensitivity of key statistics to calling effort;
3) daily data on 12 domains (2 gender groups, 2 age groups, and 3 race/ethnicity groups) that are correlated with NSFG estimates and key domains of interest;
4) data from responsive design interventions on key auxiliary variables during data collection in order to improve the balance on those variables among respondents and nonrespondents;
5) a two-phase sampling plan, selecting a probability sample of nonrespondents at the end of week 10 of each 12-week quarter;
6) data from comparisons of alternative postsurvey adjustments for nonresponse; and
7) data from a large-scale experiment with different incentives offered in phase 1.

Nonresponse bias analysis is an integral part of the design of the continuous NSFG. In addition to ongoing monitoring, we frequently conduct more detailed, specialized analyses to understand any changes in patterns of nonresponse. A detailed analysis of response rates and description of the data collection process for 2006-2010 Continuous NSFG have previously appeared (Wagner et al., 2012; Lepkowski et al., 2013). Below we describe in more detail the procedures used to monitor and manage data collection.

Results from 2011-2013 Continuous NSFG

The 2011-2013 survey used a two-phase sample design to reduce the effects of nonresponse bias, and responsive design procedures to reduce the cost of data collection. Weighted response rates overall were 73 percent among females, 72 percent among males, and 73 percent overall. These response rates are somewhat lower than those achieved in 2006-2010 NSFG Continuous, which had an overall response rate of 77%. Weighted teenage response rates for 2011-2013 were 75% for both females and males. These weighted response rates account for nonresponse to the screener and the main interview, and Phase 1 and Phase 2 nonresponse.

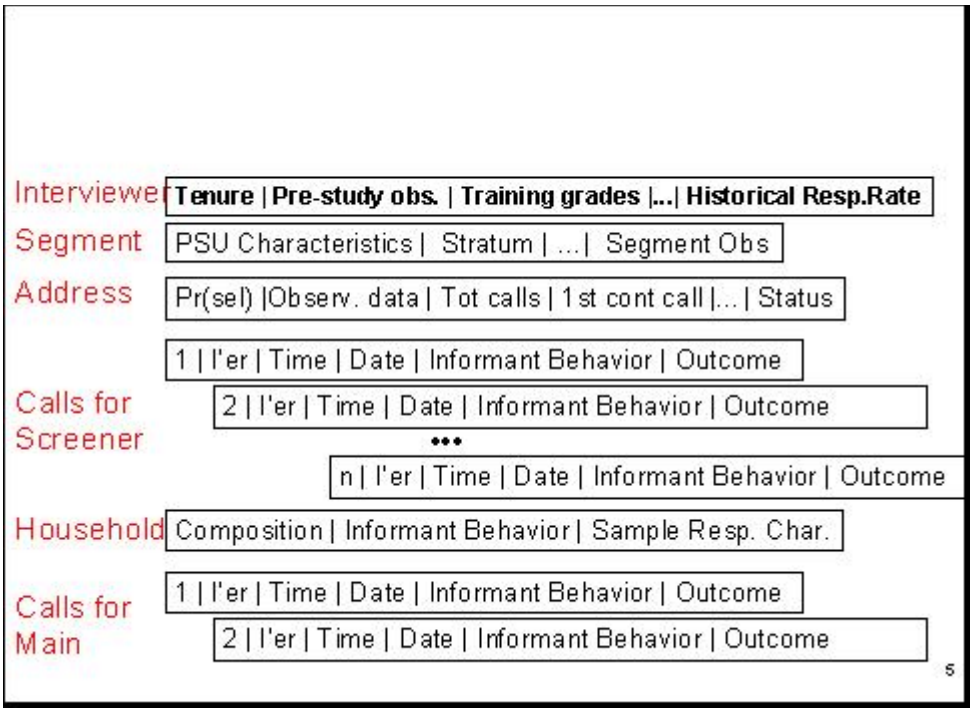
The overall screener response rate (to identify eligible persons age 15-44 for the main interview) was 93% while the main interview response rate (conditional on a completed screener) was 78%. The final weighted rates for key subgroups ranged from a low of 70 % for Other females age 20-44 to 79% for Black females ages 15-19. One of the key objectives of responsive design is to monitor the variation in these rates and intervene to minimize the differences, as one means of reducing the risk of nonresponse bias.

The balance of this report describes the key elements of the responsive design approach used during 2011-2013 and throughout the continuing 2011-2015 NSFG to manage data collection and to attempt to measure and reduce nonresponse bias.

1. Paradata Structure

The paradata for NSFG consist of observations made by listers of sample addresses when they visit segments for the first time, observations by interviewers upon first visit and each contact with the household, call record data that accumulate over the course of the data collection, and screener data about household composition.

These data can be informative about nonresponse bias to the extent that they are correlated with both <u>response propensities</u> and <u>key NSFG variables</u>. The structure of the paradata is shown below:



Thus, we have data on

(a) the interviewers,

(b) the sampled segments (including 2010 census data and segment observations by listers and interviewers),

(c) the selected address,

(d) the date and time of visits ("calls") for screeners and main interviews, and the outcomes of those visits,

(e) the sampled household, and

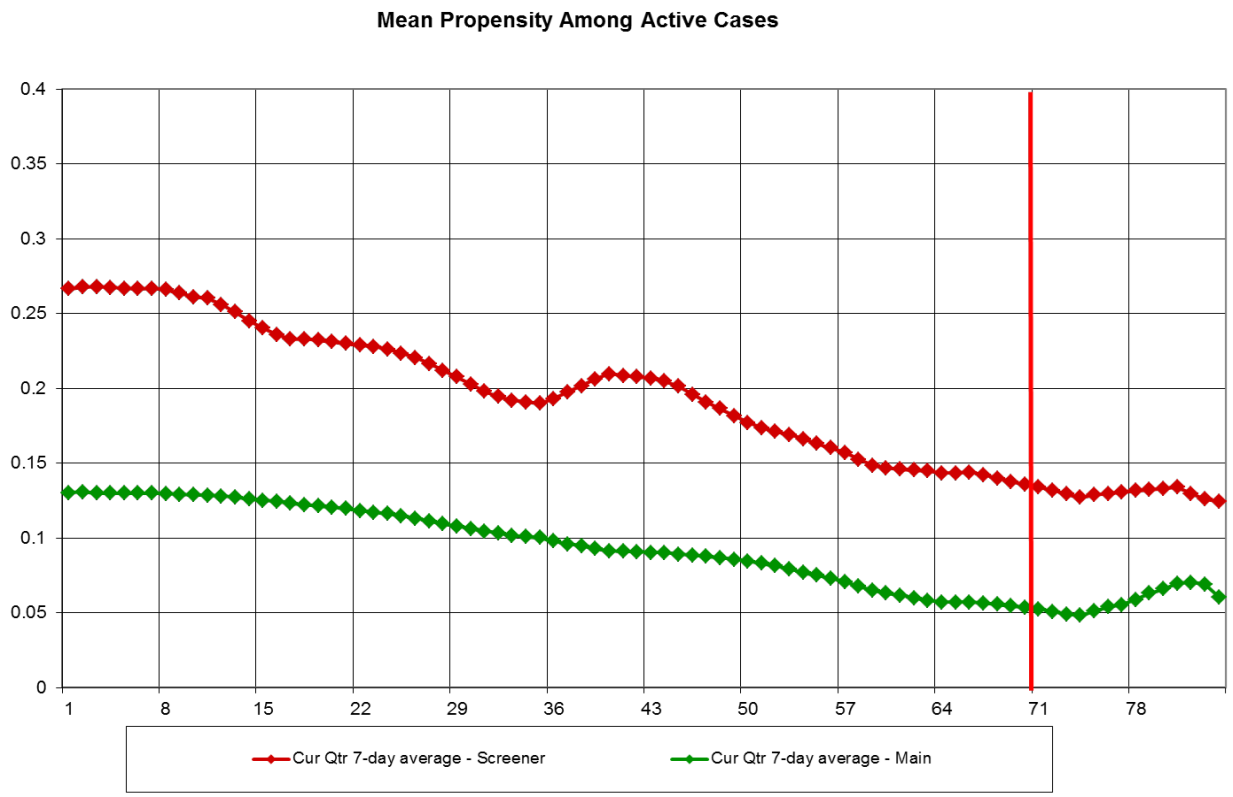
(f) for completed screeners, the selected respondent.

These data include comments or remarks made by screener informants and by persons selected to be interviewed.

From these data we build daily propensity models (using logistic regression) predicting the probability of completing an interview on the next call. These models estimate the likelihood for each active screener and main case that the next call will generate a successful interview. We monitor the mean probability of this event over the course of the 10-week Phase 1 data collection period. These data allow us to identify areas or subgroups where more effort may be needed to achieve desirable balance in the respondent data set, and intervene as necessary. These

propensity models were refit before data collection began in 2011 and cross-validated using quarterly data from 2006-2010 NSFG Continuous. Further, we considered the impact of the use of data from prior quarters in these models (Wagner and Hubbard, 2014).

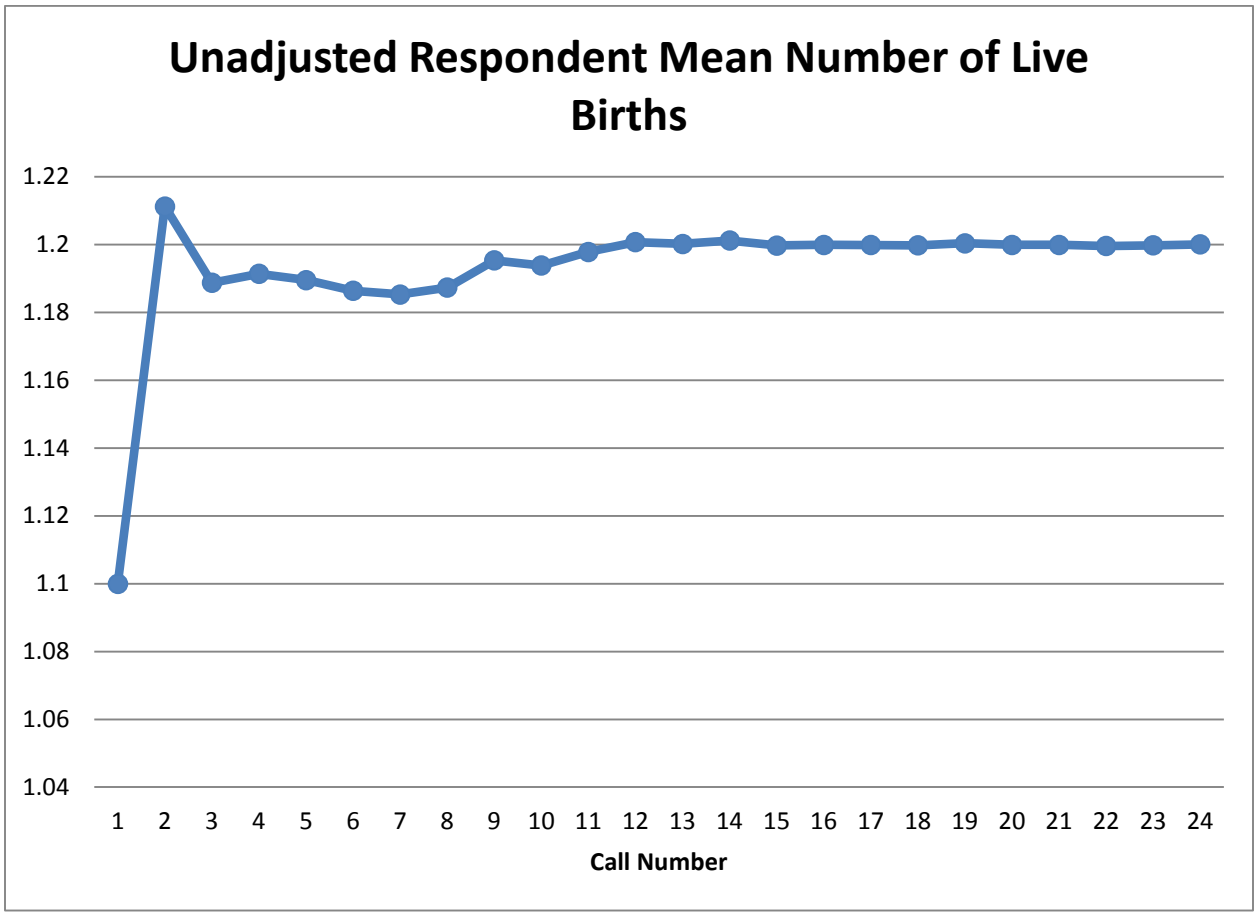
We track the mean probability of an active case responding daily throughout the data collection period, using graphs like that below. The graph below shows data for one recent 12-week (84-day) data collection period. It shows a gradual decline in the likelihood of completing an interview as the data collection period proceeds, reflecting the fact that easily accessible and highly interested persons are interviewed most easily and quickly.



2. Sensitivity of Key Estimates to Calling Effort

We estimate daily (unadjusted) respondent-based estimates of key NSFG variables. We plot these estimates as a function of the call number on which the interview was conducted, yielding graphs like that below. For example, the chart below provides the unadjusted respondent estimate of the mean number of live births among females, which stabilizes at around 1.2 within the first 9 calls. That is, the combined impact of the number of interviews brought into the data set after 9 calls and the characteristics of those cases on the "mean number of live births" variable produces no change in the unadjusted respondent estimate. For this specific measure, therefore, further calls with the phase 1 protocol have little effect.

Monitoring several of these indicators allows us guidance on minimum levels of effort that are required to yield stable results within the first phase of data collection.

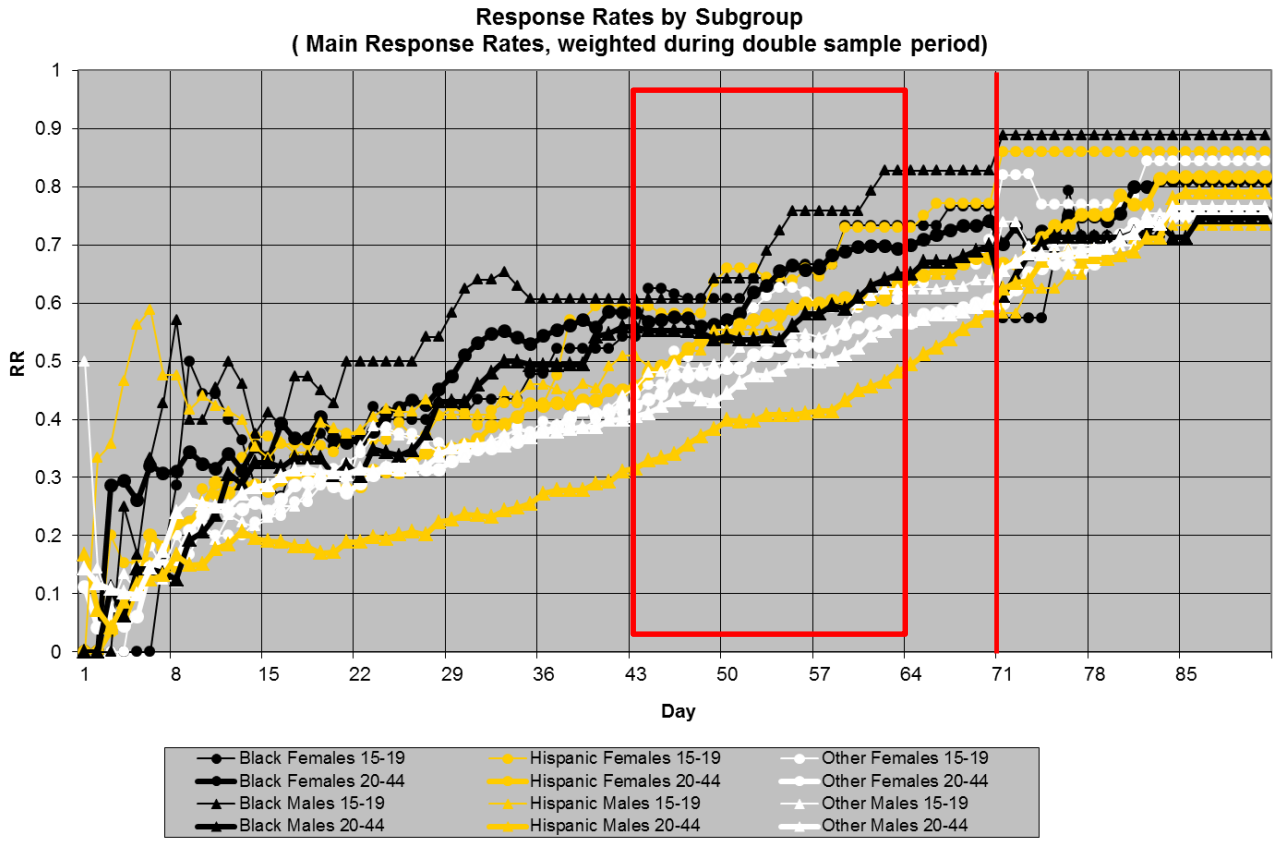


3. Daily Monitoring of Response Rates for Main Interviews across 12 Socio-Demographic Groups

We compute response rates of main interviews (conditional on obtaining a screener interview) daily for 12 socio-demographic subgroups that are domains of the sample design and important subclasses in much of family demography (i.e., 12 age by gender by race/ethnicity groups). We estimate the coefficient of variation of these response rates daily, in an attempt to reduce that variation in response rates as much as possible. When the response rates are constant across these subgroups, then we have controlled one source of nonresponse bias on many NSFG national full population estimates (that bias due to true differences across the subgroups).
In addition to monitoring these demographic subgroups, we also monitor differences in response rates based upon information obtained during the screening interview. We monitor differences in response rates between those sampled persons in households with children and without children. Further, our interviews estimate whether the sampled person is in an active sexual relationship with a person of the opposite sex. We monitor response rates for those judged to be in such a relationship and those judged not to be in such a relationship. We intervene if we see imbalances in the response rates across key subgroups or other indicators. These interventions are described in more detail below.

4. Responsive Design Interventions on Key Auxiliary Variables during Data Collection

We have conducted a variety of interventions aimed at correcting imbalances in the current respondent pool on key auxiliary variables. One example intervention from the fourth quarter of 2011-2015 NSFG Continuous is shown in the graph below. Here Hispanic male adults 20-44 (based on screener data; see the lowest yellow line) were judged to have lower response rates than the other groups at the time of the intervention (day 43 of the quarter). The intervention (shown in the red box) targeted this group for extra effort, bringing the response rate more in line with the other key demographic groups of importance to NSFG.



These interventions have been based on a variety of indicators available to us and monitored during the field period. Some examples of intervention targets include cases with addresses matched to an external database to identify households containing potentially age-eligible (or ineligible) persons; screener cases with high predicted probability of age-eligibility, based on paradata; cases with high base weights; cases with high (or low) predicted probability of response; and households with (or without) children, based on screener data. Finally, we have experimented with prioritizing cases that are predicted to be influential on key estimates (Wagner, 2014). For all these interventions, our work in 2006-2010 NSFG Continuous demonstrated with experimental evidence that these types of interventions will lead to increased interviewer effort on the prioritized cases, and this increased effort will frequently lead to increases in response rates for the targeted groups(Wagner, et al., 2012).

We have continued to refine the intervention strategies (mostly by increasing the visibility and feedback on progress for the cases sampled for the intervention), and to evaluate which types of interventions are more successful than others.

5. A Two-Phase Sampling Scheme, Selecting a Probability Sample of Nonrespondents at the End of Week 10 of Each Quarter

At the end of week 10 of each quarter, a probability subsample of remaining nonrespondent cases is selected. The sample is stratified by interviewer, screener vs. main interview status, and expected propensity to provide an interview. A different incentive protocol is applied to these cases, and greater interviewer effort is applied to the subselected cases. Early analysis of the performance of the second phase noted that outcomes for active main cases sampled into the second phase sample were better than those for screener cases; hence, the sample is disproportionately allocated to main cases (about 60% of the cases are active main interview cases). The revised incentive used in Phase 2 (weeks 11 and 12 of each 12-week quarter) appears to be effective in raising the propensities of the remaining cases, bringing into the respondent pool persons who would have remained nonrespondent without the second phase.

6. Comparison of Alternative Postsurvey Adjustments for Nonresponse

As part of the preparation of the public-use data file that includes data collected from 2011 to 2013, we developed nonresponse adjustment models. These models included variables that predicted both response and key estimates produced by the NSFG. Variables that meet these criteria are best suited for adjusting for nonresponse bias (Little and Vartivarian, 2005). Models were selected predicting both response to the screener interview and the main interview. Tables 1 and 2 below show the variables used in these models (screener and main).

Table 1. Screener response propensity predictors for nonresponse adjustment models: National Survey of Family Growth

Predictor Name	Description
REL_FAMILY_HHDS_CEN_2010	Number of 2010 Census households in the Block Group where at least 2 members are related
PLATINO_10CENSUSPL	Proportion Latino/Hispanic of Total Population from Census 2010
PBLACK_10CENSUSPL	Proportion Non-Latino/Hispanic Black of Total Population from Census 2010
URBAN	Address in an urban location (yes/no)
CALLS_CAT	Category for number of Screener Call Attempts (1=1-5; 2=6-8; 3=9)
KIDS	Interviewer Observation about Presence of Children in Housing Unit (yes/no)
PHONE_FLAG	Telephone number available for merging to address via commercial vendor
PHYSIMPED	Interviewer observed physical impediments to entry, such as locked door, community gate, etc. (yes/no)

Table 2. Main-interview, nonresponse-propensity model predictors: National Survey of Family Growth

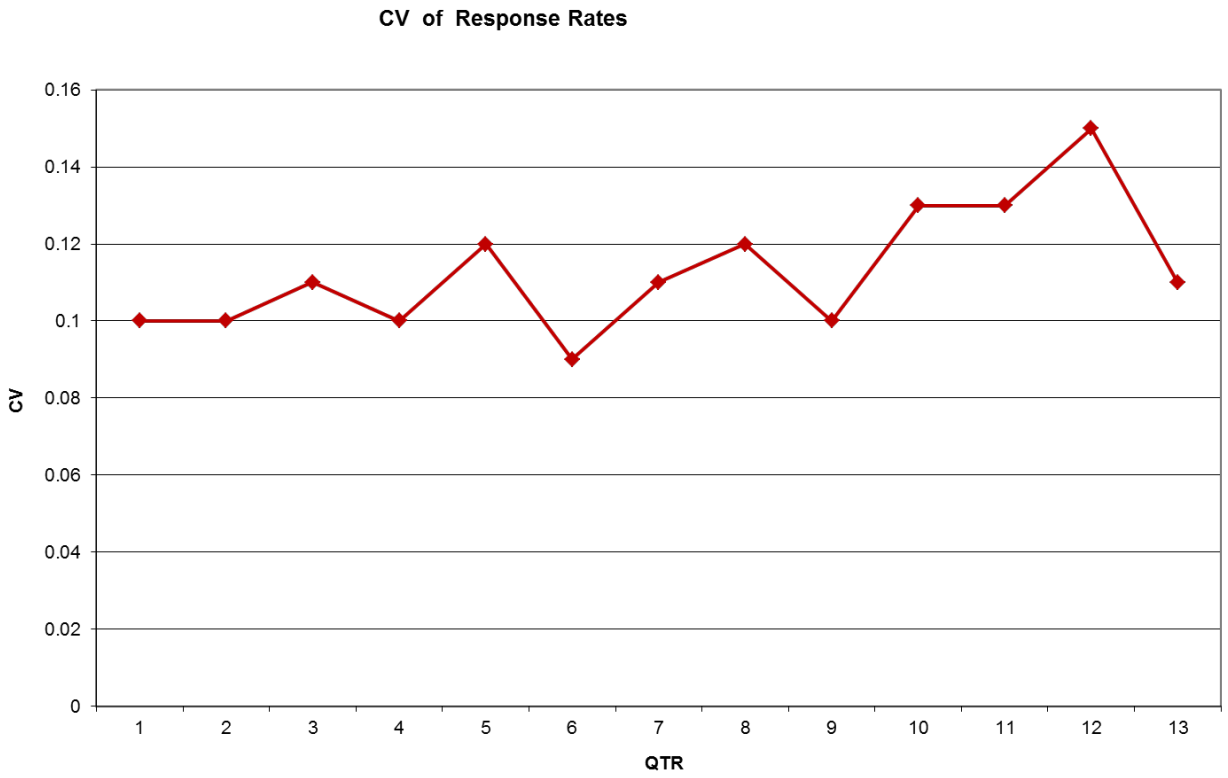
Predictor name	Predictor description
URBAN	Address in an urban location (yes/no)
MED_HOUSE_VAL_TR_ACS_06_10	Median House Value for the Census Tract Estimated from ACS 2006-2010
REL_FAMILY_HHDS_CEN_2010	Number of 2010 Census households in the Block Group where at least 2 members are related
PBLACK_10CENSUSPL	Proportion Non-Latino/Hispanic Black of Total Population
BILQ_FRMS_CEN_2010	Number of addresses that completed and returned the 2010 Census English/Spanish bilingual Mailout/Mailback form
MAIL_RETURN_RATE_CEN_2010	Number of mail returns received out of the total number of valid occupied housing units in the Mailout/Mailback universe in Census 2010 in the Block Group
ELIG_NEVER_PCT	Percentage of Census ZIP-Code Tabulation Area (ZCTA) that are Persons who have Never been Married
SEXACTIVE	Contact Obs: Respondent judged to be sexually active (yes/no)
CALLS_CAT	Category based on number of calls
SCR_SINGLEHH	Screener interview data indicate single person household (yes/on)
SCR_LANG	Screener interview data indicate anticipated interview will be in Spanish (yes/on)

Table 2 cont'd. Main-interview, nonresponse-propensity model predictors: National Survey of Family Growth

Predictor name	Predictor description
SCR_AGE	Screener interview data indicate selected response's age
SCR_HISP	Screener interview data indicate selected response is Hispanic (yes/no)
PHYSIMPED	Interviewer observed physical impediments to entry, such as locked door, community gate, etc. (yes/no)
ISAFETY_CONCERNS	Interviewer noted safety concerns about segment during segment listing or updating procedure (yes/no)
BLNON_ENGLISH_LANG _SPANISH	Segment Obs: Evidence of Spanish Speakers

Nonresponse adjustments were created by forming deciles of estimated propensities from each model (screener and main) separately, and then using the inverse of the response rate within each decile as an adjustment weight. The product of the screener and main nonresponse adjustments was then multiplied by the probability of selection weight to obtain the final weight.

One of the positive impacts of the interventions during 2006-2010 NSFG Continuous was a reduction in the variation of response rates for important subgroups. These relatively low rates of variation have been largely maintained in the 2011-2015 Continuous NSFG. The graph below shows how the coefficient of variation over the first 13 quarters of the 2011-2015 Continuous NSFG. The subgroup response rates are for the 12 cells defined by the cross-classification of 2 genders, 2 age groups (15-19 and 20-44), and 3 race/ethnicity groups (Black, Hispanic, White/Other). To the extent that these factors relate to survey outcomes, reducing the variation of subgroup response rates should reduce the nonresponse bias of unadjusted means estimated from the survey data. Improving response for groups with relatively lower response rates is also an empirical test of the assumption that within subgroups, responders are a random sample of the sample.



7. Incentive Experiment

In response to the decline in response rates and rising costs relative to 2006-2010 NSFG Continuous, a change in the incentive structure for the NSFG was proposed. The experiment and its results have been extensively reported in separate communication with OMB. Here, we note only the main results.

The experiment changed the Phase 1 incentive from $40 to $60. The Phase 2 incentive was left at $80 for all respondents. The experiment ran for five quarters. We found that the $60 incentive increased response rates in Phase 1 and improved the number of interviews obtained, but that these gains were lost during Phase 2 as the $40 incentive group had similar response rates at the end of Phase 2. The $60 incentive increased the efficiency of interviewing, but only by an amount roughly equivalent to the cost of the added incentive. We also found that the larger incentive did not lead to interviewing persons who differed in terms of the key estimates. As a result of these findings, it was decided that we would continue with the $40 incentive in Phase 1.

Summary

In summary, the 2011-2015 Continuous NSFG builds on the design and implementation of the 2006-2010 Continuous NSFG. A key element of that design is a responsive design approach that monitors paradata and key statistics, with a view to minimizing nonresponse bias and maximizing field efficiency. The NSFG team continues to explore new design options aimed at improving response rates, the composition of response, and efficiency.

References for this document

Groves, R. M., W. D. Mosher, J. M. Lepkowski and N. G. Kirgis (2009). "Planning and Development of the Continuous National Survey of Family Growth." Vital Health Stat 1(48): 1-64.

Lepkowski, J.M., Mosher, W.D., Davis, K.E., Groves, R.M., and Van Hoewyk, J. 2010. The 2006-2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey. Vital and Health Statistics, Series 2, No. 150. Hyattsville, MD: National Center for Health Statistics.

Lepkowski, J. M., W. D. Mosher, R. M. Groves, B. T. West, J. Wagner and H. Gu (2013). Responsive Design, Weighting, and Variance Estimation in the 2006-2010 National Survey of Family Growth, National Center for Health Statistics. Series 2. Hyattsville, MD: National Center for Health Statistics.

Little, R. J. A. and S. Vartivarian (2005). "Does Weighting for Nonresponse Increase the Variance of Survey Means?" Survey Methodology **31**(2): 161-168.

Wagner, J. (2014). Limiting the Risk of Nonresponse Bias by Using Regression Diagnostics as a Guide to Data Collection. Conference Proceeding of the Joint Statistical Meetings, Boston.

Wagner, J. and F. Hubbard (2014). "Producing Unbiased Estimates of Propensity Models During Data Collection." Journal of Survey Statistics and Methodology **2**(3): 323-342.

Wagner, J., B. T. West, N. Kirgis, J. M. Lepkowski, W. G. Axinn and S. K. Ndiaye (2012). "Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection." Journal of Official Statistics **28**(4): 477-499.