Nonresponse Bias Analyses for the
Continuous National Survey of Family Growth (NSFG), 2011-2019
Prepared for the NCHS/NSFG Team by
James Wagner, Ph.D. NSFG Chief Mathematical Statistician,
and Mick P. Couper, Ph.D., NSFG Project Director,
University of Michigan

Outline:

Executive Summary

This brief appendix describes our approach to nonresponse bias analysis and fieldwork management on the National Survey of Family Growth (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 summaries for 2011-2013, 2013-2015, 2015-2017, and 2017-2019 on the NCHS' website) 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 2011-2019 period of the Continuous NFSG, that is, all 8 years of fieldwork (2011-2019) covered under the current NSFG contract. These activities have built upon the success of the 2006-2010 NSFG, the first survey period in which NSFG used a continuous design, and have made use of further 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 Continuous NSFG 2011-2019 experienced response rates below the 80% target set by OMB. The fieldwork for the NSFG 2011-2019 was completed in August 2019. NCHS has released four sets of public-use data files based on data collected from 2011-2013, 2013-2015, 2015-2017, and 2017-2019. In this report, we review results from 2011-2019 (i.e. the period corresponding to the four public use files combined).

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) Response rates for the 2011-2019 period corresponding to four public use files,
2) quarterly measures of effort (e.g. mean calls per completed interview) and response rate

for Quarter 1 to Quarter 32 (Q1 to Q32);

3) a paradata structure that uses lister and interviewer observations of attributes related to response propensity and some key survey variables;

4) data on the sensitivity of key statistics to calling effort;

5) 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;

6) 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;

7) a two-phase sampling plan, selecting a probability sample of nonrespondents at the end of week 10 of each 12-week quarter; and

8) results from models developed in the context of postsurvey adjustments for nonresponse.

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 the 2006-2010 NSFG, with a continuous fieldwork design, have previously appeared (Wagner et al., 2012; Lepkowski et al., 2013; and summaries for 2011-2013, 2013-2015, 2015-2017, and 2017-2019 on the NCHS' website). Below we describe in more detail the procedures used to monitor and manage data collection.

1. 2011-2019 Response Rates from the Continuous NSFG (Q1-Q32; Years 1-8)

NSFG fieldwork in 2011-2019 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 for this 8-year span from September 2011 to August 2019 were 69.0% among females, 65.8% among males, and 67.5% overall. These response rates are lower than those achieved in the 2006-2010 NSFG, which had an overall response rate of 77%. This decline is consistent with the trend facing all major surveys in the U.S (Williams and Brick, 2017). Weighted response rates for teens 15-19 in 2011-2019 were 69.8% for females and 68.9% for males. These weighted response rates account for nonresponse to the screener and the main interview, and Phase 1 and Phase 2 nonresponse. The 2011-2019 screener response rate (to

identify eligible persons age 15-49 [or 15-44 for 2011 to 2015] for the main interview) was 89.6% while the main interview response rate (conditional on a completed screener) was 75.4%.

The balance of this report describes the key elements of the responsive design approach used in 2011-2019 for the Continuous NSFG to manage data collection and to attempt to measure and reduce nonresponse bias across those 32 quarters of data collection.

2. Quarterly Measures of Effort and Response Rates for Q1 to Q32

We see consistent trends over the time period from Q1 through Q32. Figure 1 shows the Phase 1 and Phase 2 response rates by quarter. The Phase 1 protocol has a diminishing effectiveness over this time period. Phase 2, on the other hand, remains fairly consistent over time in its impact, but is unable to make up for the diminishing impact of Phase 1.

Figure 1. Phase 1 and Final Response Rates by Quarter (Q1-Q32)
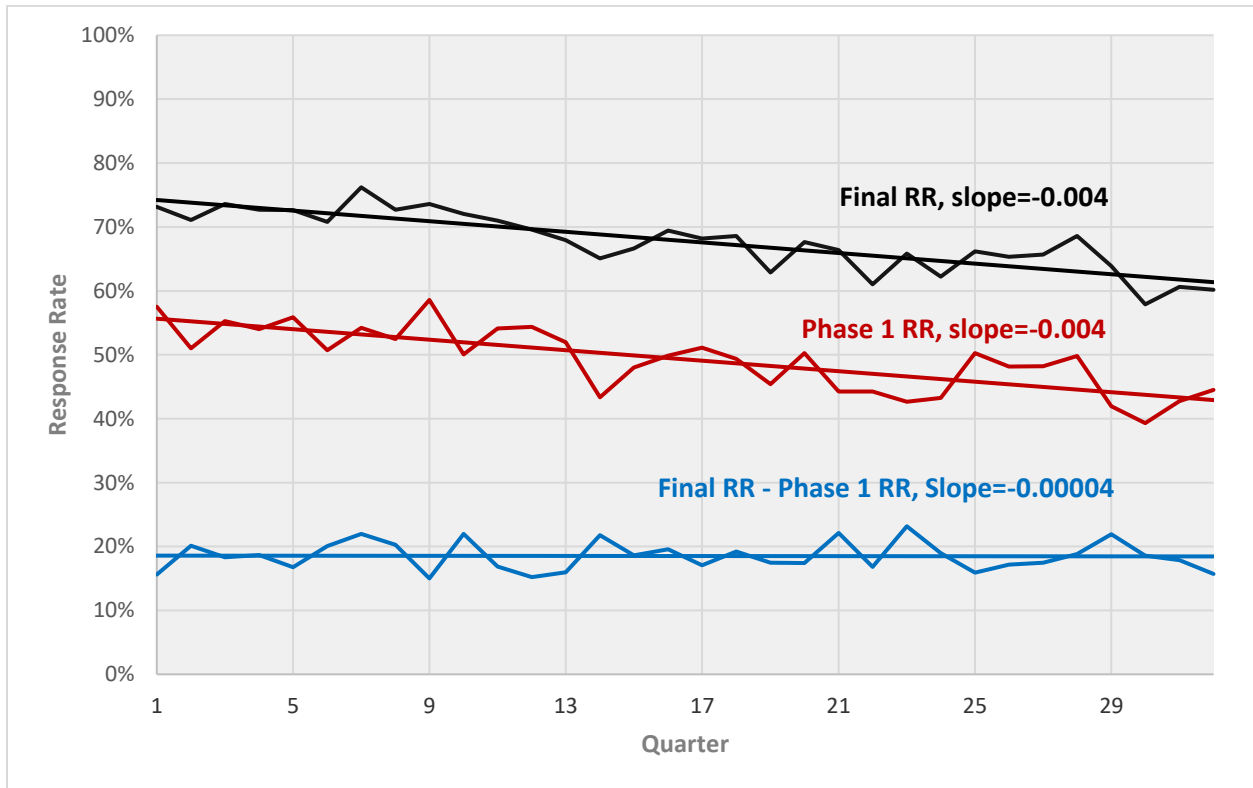
Over this time period, we have observed that completing interviews required more effort. Contact rates declined (see Figure 2), and the number of calls required to complete an interview increased (see Figure 3).

Figure 2. Call-Level Contact Rates for Screener and Main Interviewing by Quarter (Q1-Q32)
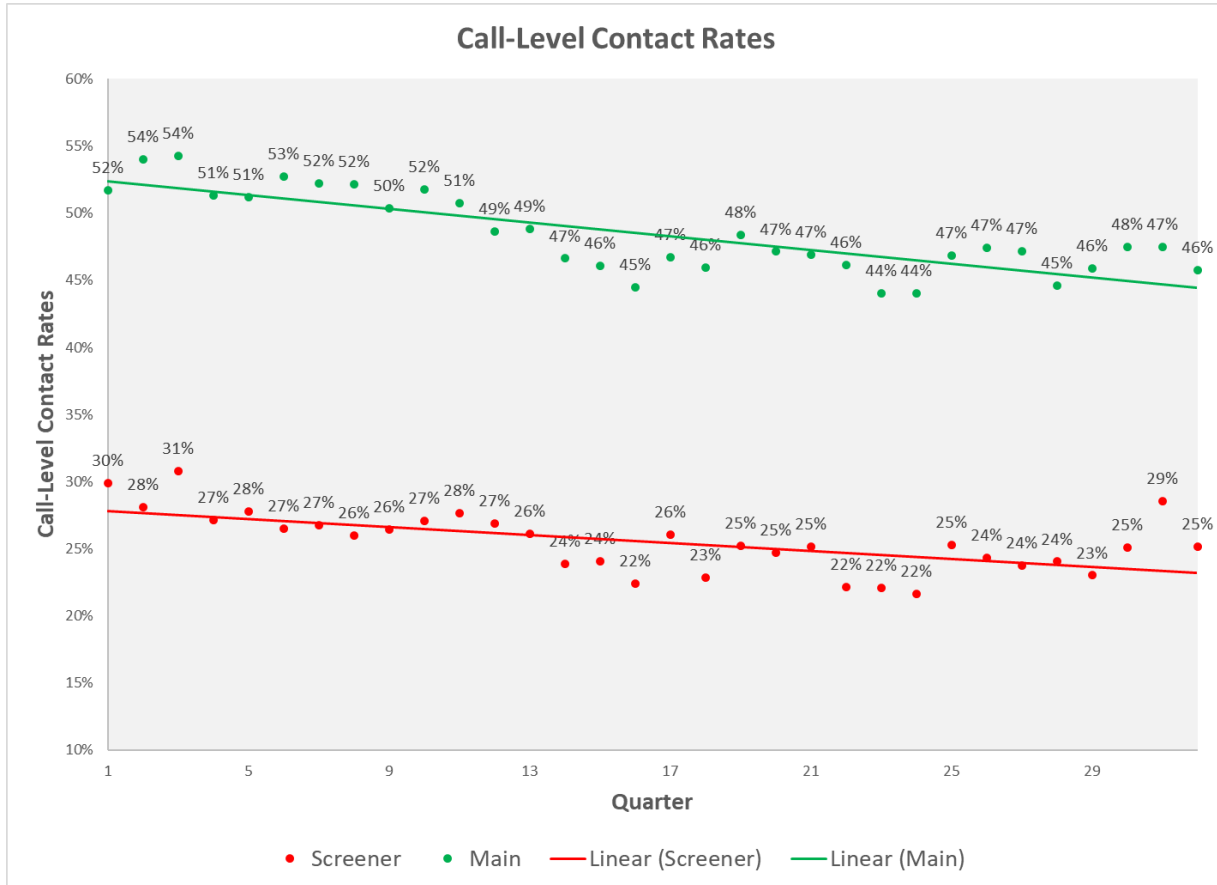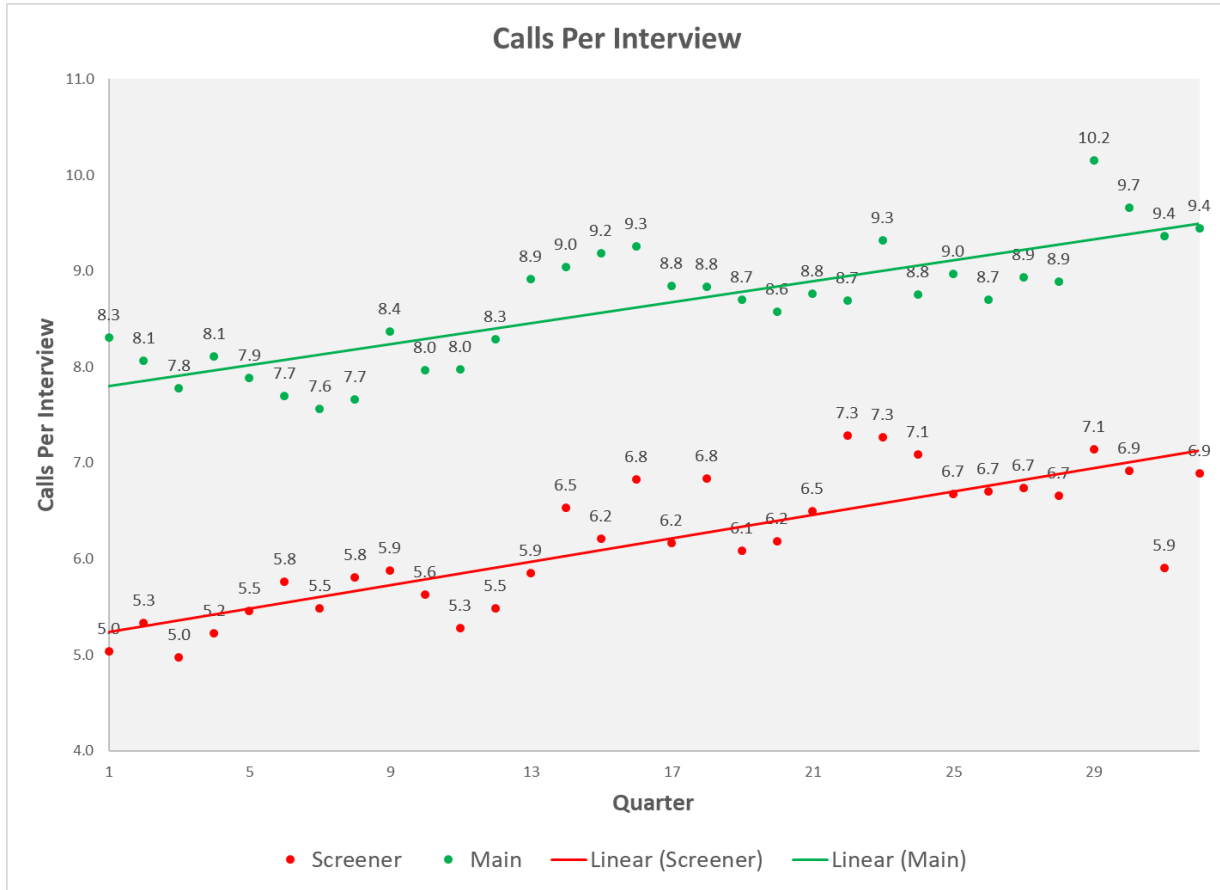
Figure 3. Mean Calls Per Interview for Screener and Main Interviewing by Quarter (Q1-Q32)

3. 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 response propensities and key NSFG variables. The structure of the paradata is shown below in Figure 4:

Figure 4. NSFG Paradata Structure
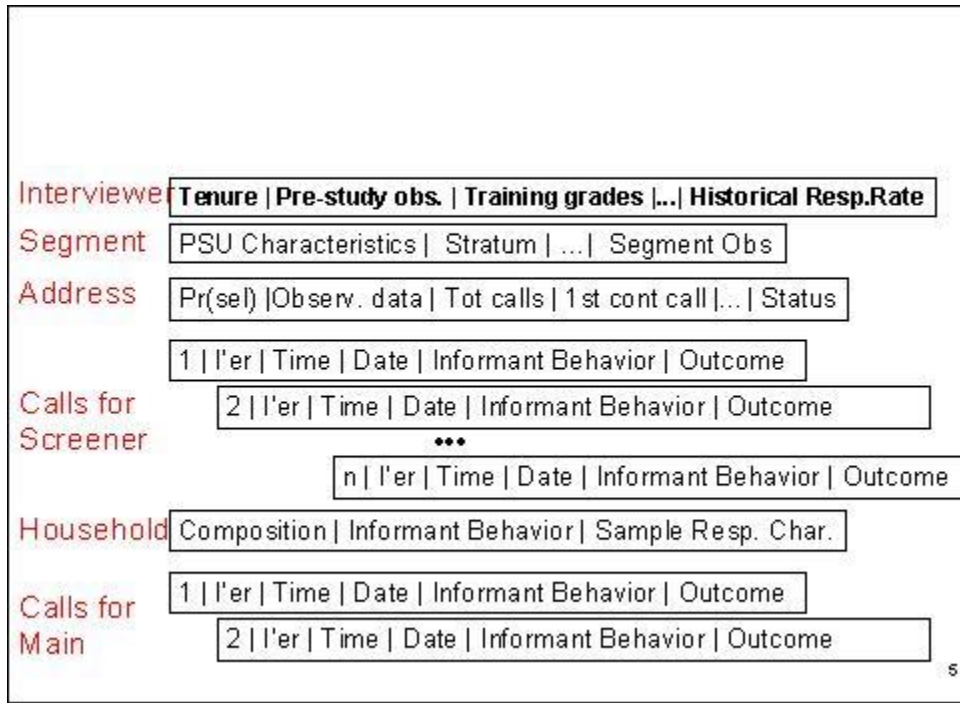
Thus, we have data on

(a) the interviewers,

(b) the sampled segments (including 2010 Decennial Census data, data from the continuous American Community Survey, data from the Census Bureau's Planning Database (e.g. mail response rate to 2010 Decennial Census in the Census Tract) 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 the 2006-2010 NSFG. Further, we considered the impact of the use of data from prior quarters in these models (Wagner and Hubbard, 2014). We have also been developing Bayesian approaches to the estimation of these models where data from previous quarters, literature reviews, and expert opinions are used as prior information (Wagner, 2016; Wagner 2017; West, et al., 2019; Coffey, et al., 2020).

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. Data from the same quarter (i.e. the same season, such as Sep-Dec) of previous years are also presented for comparison. The red vertical line indicates the phase boundary between phases 1 and 2.

Figure 5. Mean Propensity among Active Cases

**Mean Propensity Among Active Cases**
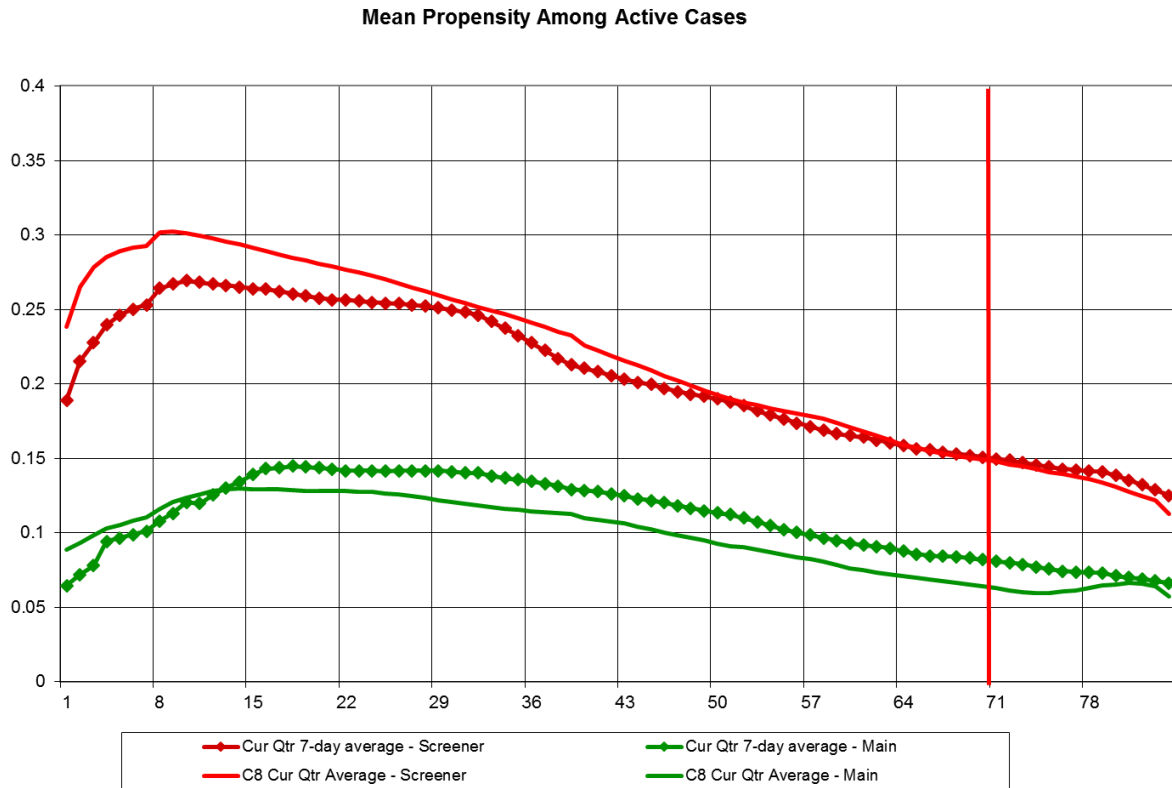
4. Sensitivity of Key Estimates to Calling Effort

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 just below 1.2 within the first 6 calls (compare values to scale on the left). That is, the combined impact of the number of interviews brought into the data set after 6 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. A second variable, number of sexual partners in the last 12 months, is also presented. This key statistic stabilizes at just less than 1.2, also around call number 6.

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.

Figure 6. Two Key Statistics by Call Number (Q1-Q32)
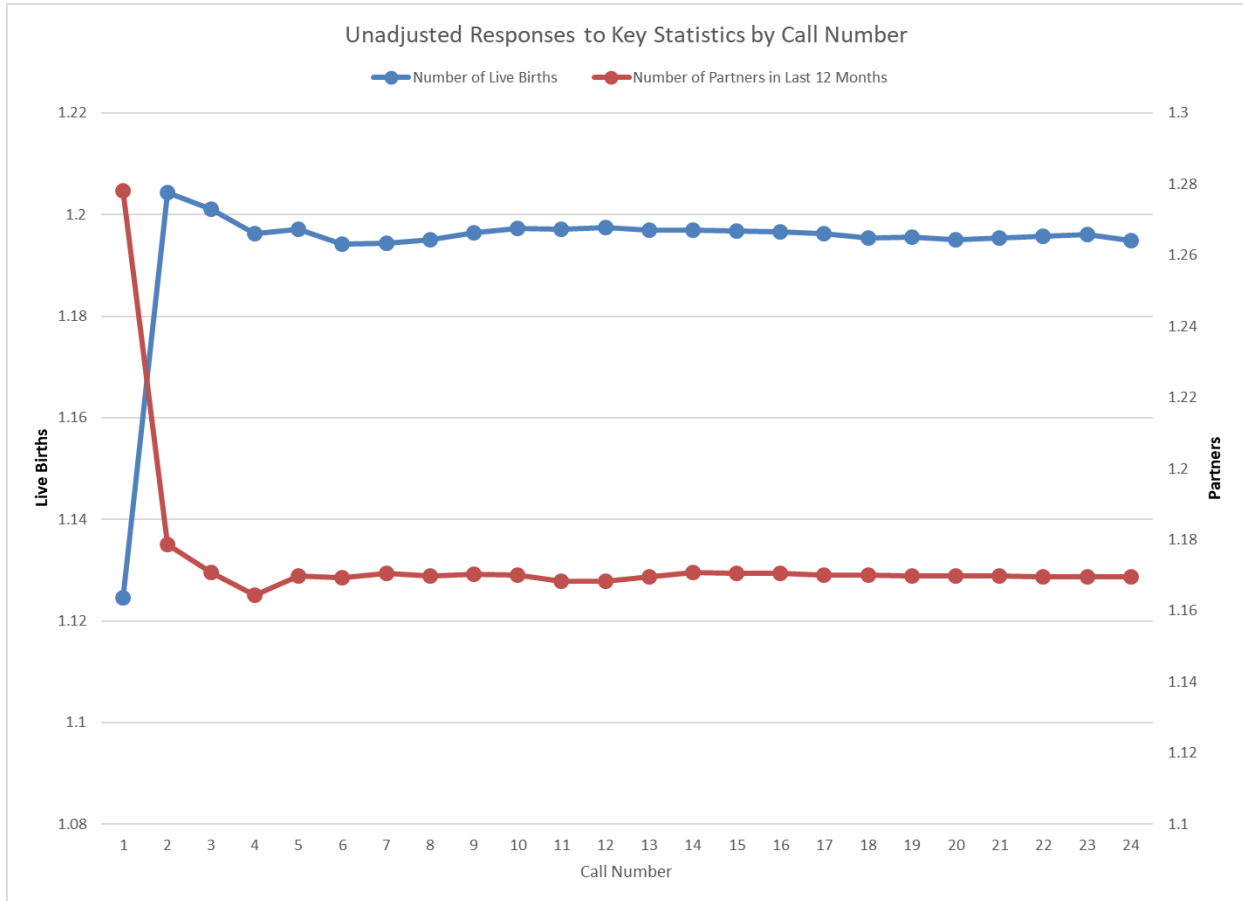
5. Daily Monitoring of Response Rates for Main Interviews across 12 Socio-Demographic Groups

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. We published a paper demonstrating that correcting these imbalances on observed variables tends to reduce the nonresponse bias of adjusted estimates from the survey data (Schouten, et al., 2016).

6. 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 Q20 of the Continuous NSFG is shown in the graph below. Here Hispanic male adults 20-49 (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.

Figure 7. Daily Subgroup Response Rates with Intervention Period Highlighted

Response Rates by Subgroup
(Main Response Rates, weighted during double sample period)
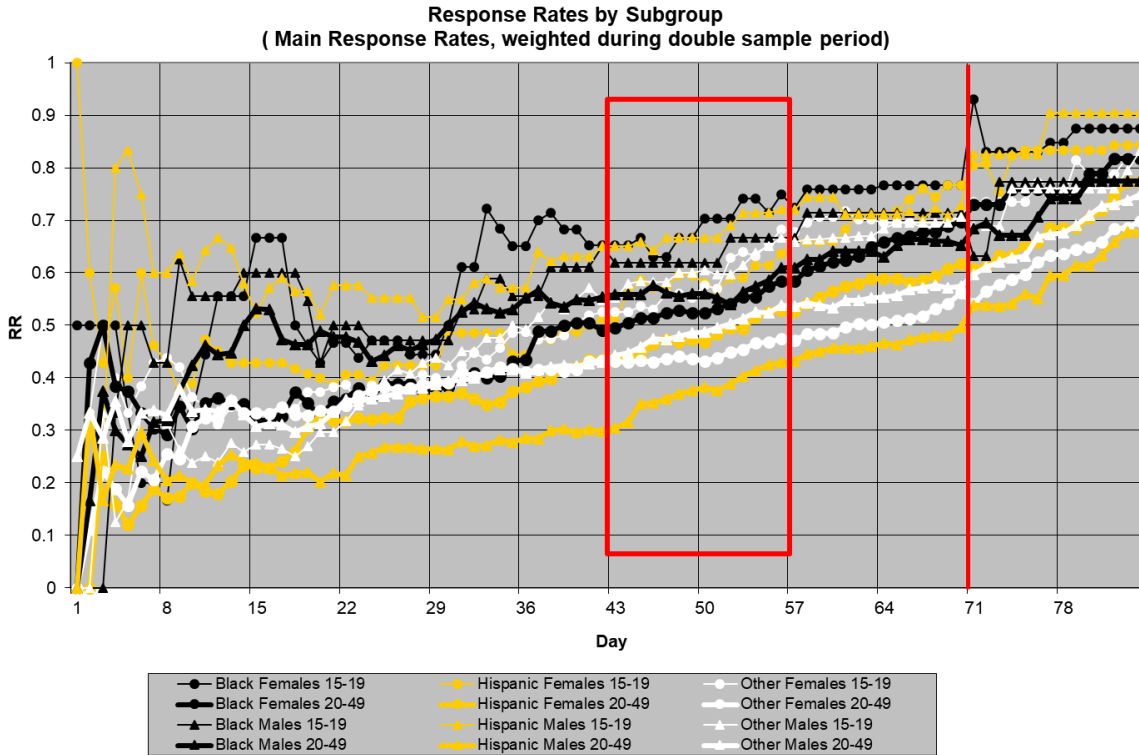
These interventions have been based on a variety of indicators available to us and monitored during the field period. Some examples of intervention targets include cases with addresses matched to an external database to identify households containing potentially eligible (or ineligible) persons; screener cases with high predicted probability of 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 the 2006-2010 NSFG 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). The following table shows the subgroup response rates for various demographic subgroups from the 32 quarters of data collection.

Table 1. Subgroup Response Rates for Q1-Q32

	Interviews	Weighted Response Rates (95% Conf Interval)
Male	**19,067**	**65.8%(65%-66%)**
15-19	4,041	68.9%(68%-70%)
20-49*	15,026	65.1%(65%-66%)
Black	3,705	71.9%(71%-73%)
Hispanic	4,252	65.0%(64%-66%)
White/Other	11,110	64.4%(64%-65%)
Female	**22,995**	**69.0%(68%-69%)**
15-19	3,971	69.8%(69%-71%)
20-49*	19,024	68.8%(68%-69%)
Black	5,163	73.8%(73%-75%)
Hispanic	5,453	70.4%(69%-71%)
White/Other	12,379	66.6%(66%-67%)
TOTAL	**42,062**	**67.5%(67%-68%)**

*20-44 for 2011-2015

We continued to refine the intervention strategies over time (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.

7. 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 active main cases are oversampled, resulting in a second phase sample that has a higher proportion of main lines than the sample that is active at the end of Phase 1. 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.

8. Results from Models Developed in the Context of Postsurvey Adjustments for Nonresponse

As part of the preparation of the public-use data file that includes data collected from September 2017 to August 2019, 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 2 and 3 below show the variables used in these models (screener and main).

Table 2. Screener response propensity predictors for nonresponse adjustment models: NSFG 2017-2019

Predictor Name	Description
MANYUNITS	Interviewer observed characteristic of housing units. Categories include single family home vs multi-unit dwelling.
BASE_WEIGHT	Probability of selection weight including housing unit selection probability, second phase selection probability, and within household selection probability
EST_ELIG_RATE	Estimated age eligibility rate for the ZIP Code Tabulation Area using Census 2010 data
MED_HHD_INC_TR_ACS_12_16	Median House Income for the Census Tract estimated from ACS 2012-2016
SNGL_PRNS_HHD_CEN_2010	Number of 2010 Census households where a householder lives alone in the Block Group
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
MSG_NUMBEROFCHILDREN	An estimated count of the number of children in the household. Based upon commercial data. Treated as categorical data with a category for missing.
MSG_AGE	Estimated age of the householder. Based upon commercial data. Treated as categorical data with a category for missing.
MSG_P2AGECALC	Estimated age of "second" adult. Based upon commercial data. Treated as categorical data with a category for missing.
MSG_BESTMATCH	A measure provided by the vendor of the likelihood that the match is accurate. A categorical variable with a category for missing.

Table 3. Main-interview, nonresponse-propensity model predictors: NSFG 2017-2019

Predictor name	Predictor description
MED_HHD_INC_TR_ACS_12_16	Median household income for the Census Tract estimated from ACS 2012-2016
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
PBLACK_10CENSUSPL	Percent of population in the Census Block that is Non-Hispanic Black from Decennial Census 2010.
POTHER_10CENSUSPL	Percent of population in Census Block that is neither Non-Hispanic Black, nor Hispanic from Decennial Census 2010.
HISP_PERC	Percent of population in Census Block Group that is Hispanic from Decennial Census 2010.
EST_ELIG_RATE	Estimated age eligibility rate for the ZIP Code Tabulation Area using Census 2010 data
ELIG_NEVER_PCT	Estimated percentage of age-eligible population (15-44) that has never been married for the ZIP Code Tabulation Area using Census 2010 data
MANYUNITS	Interviewer observed characteristic of housing units. Categories include single family home vs multi-unit dwelling.
NUMKIDS14	Screening interview count of number of children 14 years of age and under in the household
SEXACTIVE	Contact Obs: Respondent Sex Active (yes/no)
SCR_HISP	Screener interview data indicate selected response is Hispanic (yes/no)
SCR_RACE	Screener interview data indicate selected respondent is Black (yes/no)
SCR_TEEN	Screener interview data indicate selected respondent is teen, i.e. age 15-19 (yes/no)
SCR_SINGLEHH	Screener interviewer data indicated selected respondent is only adult in the household (yes/no).
WITHIN_HHPROB	The within-household selection probability of the selected respondent.

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 and then poststratified to obtain the final weight.

One of the positive impacts of the interventions experimentally tested during the 2006-2010 NSFG was a reduction in the variation of response rates for important subgroups. These relatively low rates of variation have been – despite our efforts – increasing slightly over time. Figure 8 shows the coefficient of variation from Q1 to Q32. 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). In Q17 and following, the age eligibility was extended to 49. For Q17 to Q32, the age groups are 15-19 and 20-49. 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.

Figure 8. Coefficient of Variation of 12 Subgroup Response Rates by Quarter Q1-Q32
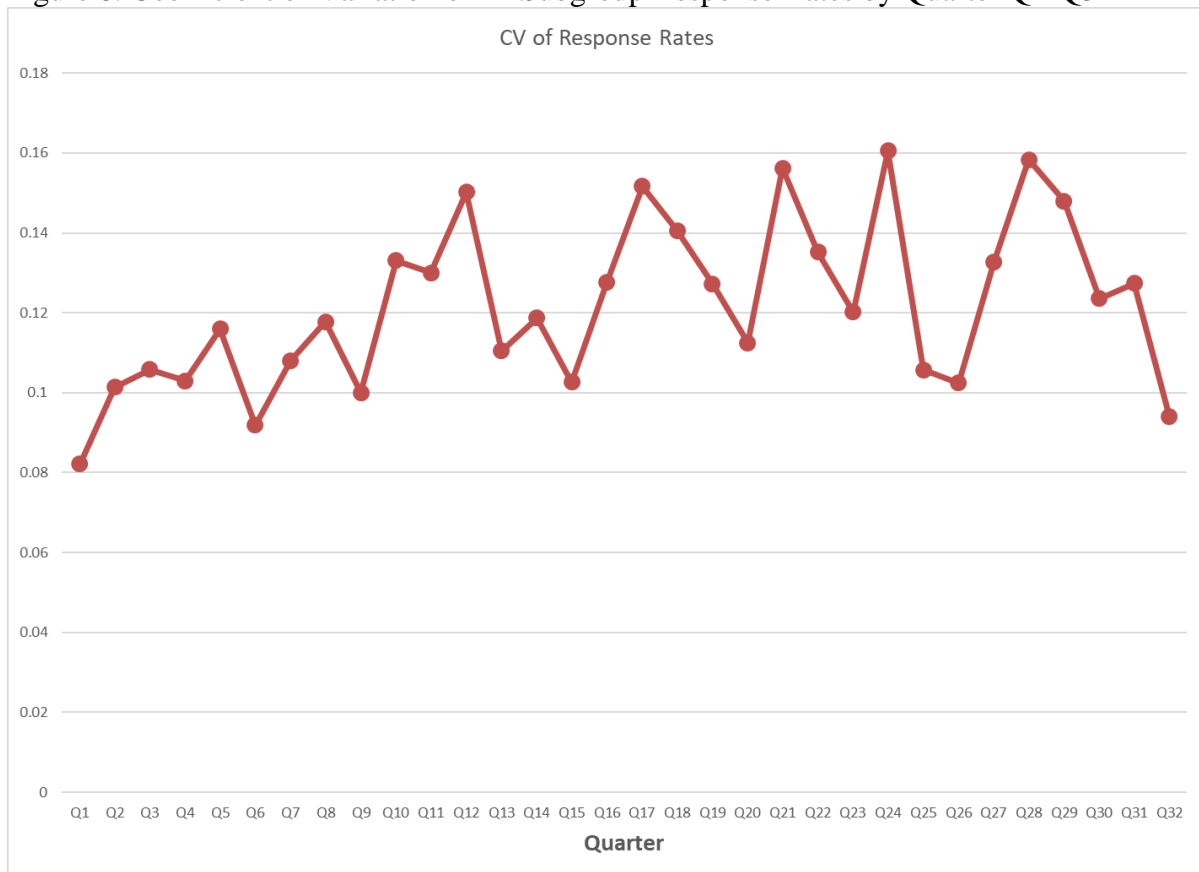
Summary

In summary, the NSFG has continued to build on the success of the earlier continuous implementation, identifying further enhancements and expanding the key metrics being monitored. The quarterly continuous data collection and two-phase design have the explicit goal of minimizing the potential for nonresponse bias while maximizing field efficiency.

References

Coffey, S., B. T. West, J. Wagner and M. R. Elliott (2020). "What Do You Think? Using Expert Opinion to Improve Predictions of Response Propensity under a Bayesian Framework." Methods, Data, Analyses **14**(2): 36.

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., 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.

Schouten, B., F. Cobben, P. Lundquist and J. Wagner (2016). "Does more balanced survey response imply less non-response bias?" Journal of the Royal Statistical Society: Series A (Statistics in Society) **179**(3): 727-748.

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

Wagner, J. (2016). Using Bayesian Methods to Estimate Response Propensity Models During Data Collection. Paper presented at the Annual Conference of the American Association for Public Opinion Research.

Wagner, J. (2017). Using Bayesian Methods to Rank Cases Based on Response Propensity During Data Collection. Paper presented at the Survey Research Methods Section of Joint Statistical Meetings.

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.

West, B.T., Wagner, J., Coffey, S., and Elliott, M.R. (2019). The Elicitation of Prior Distributions for Bayesian Responsive Survey Design: Historical Data Analysis vs. Literature Review. Retrieved from https://arxiv.org/ftp/arxiv/papers/1907/1907.06560.pdf.

Williams, D. and J. M. Brick (2017). "Trends in Us Face-to-Face Household Survey Nonresponse and Level of Effort." Journal of Survey Statistics and Methodology **6**(2): 186-211.