

Attachment N:

Nonresponse Bias Analyses for the Continuous NSFG, 2011-2015
By Mick P. Couper, Ph.D., NSFG Project Director,
and James Wagner, Ph.D. NSFG Senior Mathematical Statistician,
University of Michigan

-

Executive Summary

This brief appendix describes our approach to nonresponse bias analysis and management on the NSFG. Further details will be published in 2012 in an NCHS report (Lepkowski et al., 2012a). 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 recently-completed 2006-2010 Continuous NSFG. The 2011-2015 Continuous NSFG builds on this success, 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. Given that the design of the 2011-2015 Continuous NSFG is essentially identical to that from the previous rounds of continuous interviewing (2006-2010

NSFG), we have good prior information on likely response rates. Further, the procedures developed previously in NSFG to measure and control nonresponse bias will be used for the new round of data collection beginning in Fall 2011. We continue to develop and refine these measures, but in this attachment, we report on the activities undertaken in 2006-10 to analyze 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 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 strongly correlated with NSFG estimates;
- 4) data from randomized 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; and
- 6) data from comparisons of alternative postsurvey adjustments for nonresponse.

Nonresponse bias analysis is an integral part of the design of the continuous NSFG. At this writing, detailed analysis of nonresponse bias is underway to inform the 2011-2015 Continuous NSFG. A detailed analysis of response rates and description of the data collection process will

appear in 2012 (Lepkowski et al., 2012a), and we are currently working on more detailed assessments of nonresponse bias in the 2006-2010 Continuous NSFG (Lepkowski et al., 2012b; Wagner et al., 2012). Below we describe in more detail the procedures used to monitor and manage data collection.

Results from 2006-2010 Continuous NSFG

The 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 78 percent among females, 75 percent among males, and 77 percent overall. Weighted teenage response rates were also 77% for both females and males. This weighted response rate accounts for nonresponse to the screener and the main interview, and Phase 1 and Phase 2 nonresponse.

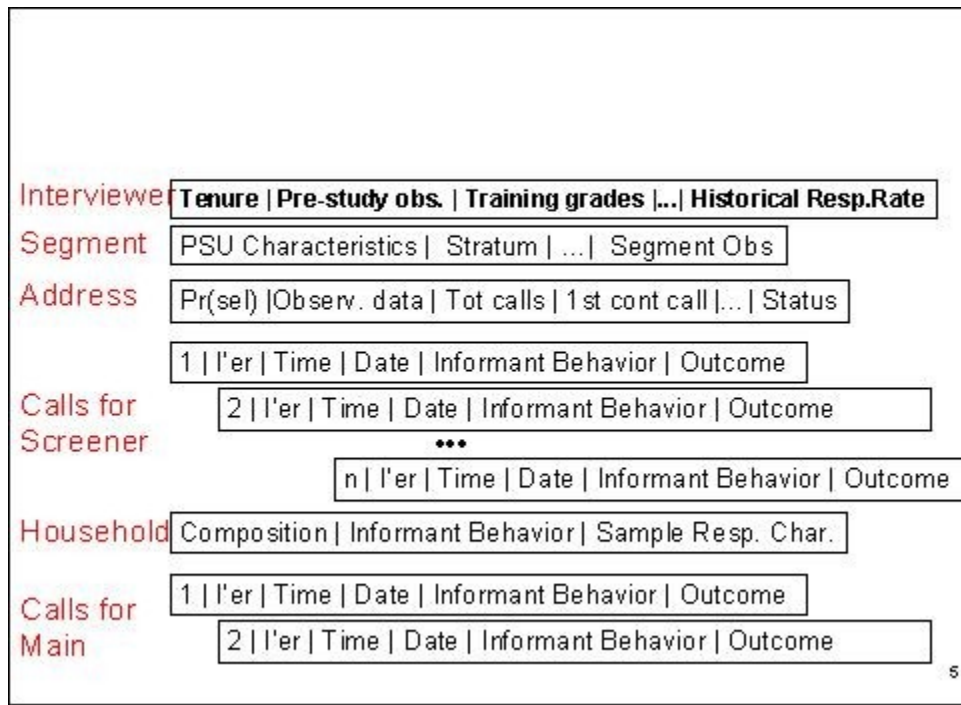
The overall screener response rate (to identify eligible persons age 15-44 for the main interview) was 93% while the main interview response rate (conditional on a completed screener) was 82%. The final weighted rates for key subgroups ranged from a low of 72.9% for Hispanic males age 20-44 to 81.9% 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 in 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 response propensities and key NSFG variables. The structure of the paradata is shown below:



Thus, we have data on

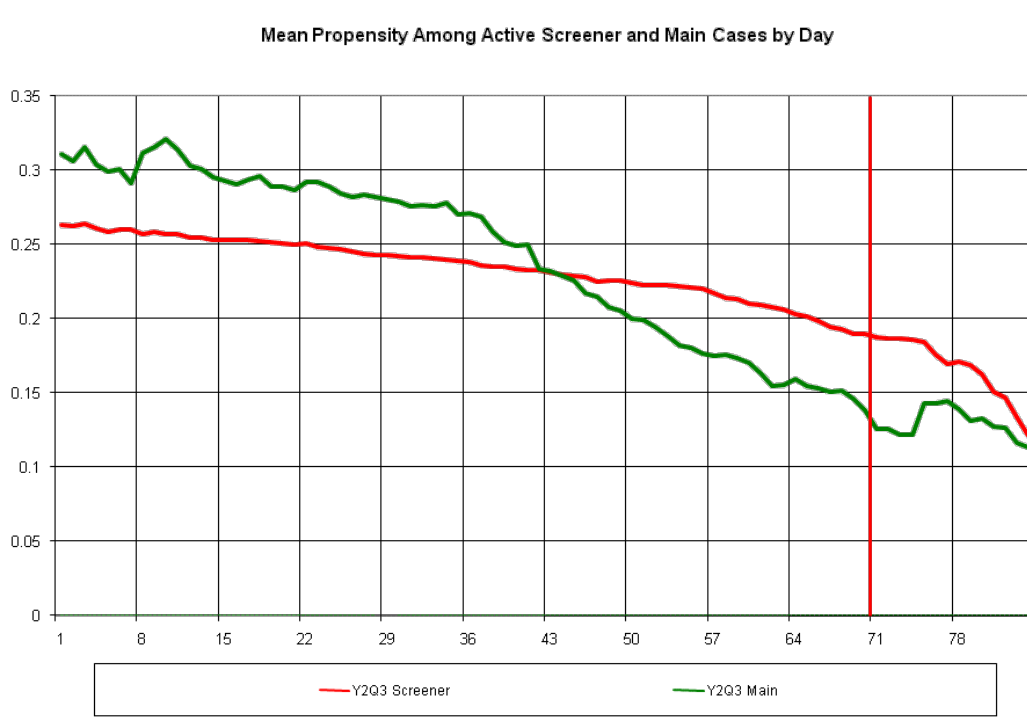
- (a) the interviewers,
- (b) the sampled segments (including 2010 census data and segment observations by listers and interviewers),
- (c) the selected address,
- (d) the date and time of visits (“calls”) for screeners and main interviews, and the outcomes of those visits,
- (e) the sampled household, 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. The specification of these propensity models is described in more detail in the forthcoming Series 2 report (Lepkowski et al., 2012a).

We track the mean propensity of an active case responding daily throughout the data collection period, using graphs like that below. The graph below shows data for one 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.

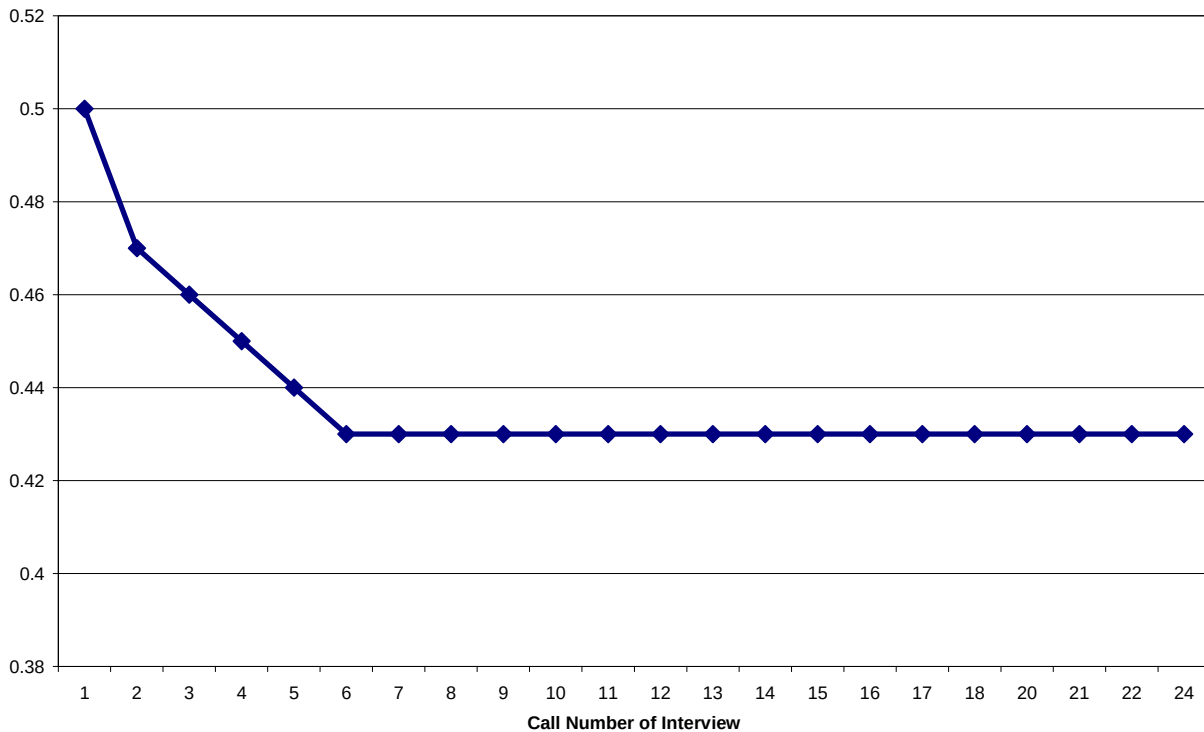


2. Sensitivity of Key Estimates to Calling Effort

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 proportion of females never married, which stabilizes at around 43% within the first 7 calls. That is, the combined impact of the number of interviews brought into the data set after 7 calls and the characteristics of those cases on the “ever married” 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.

Unadjusted Respondent Proportion Never Married (Female)



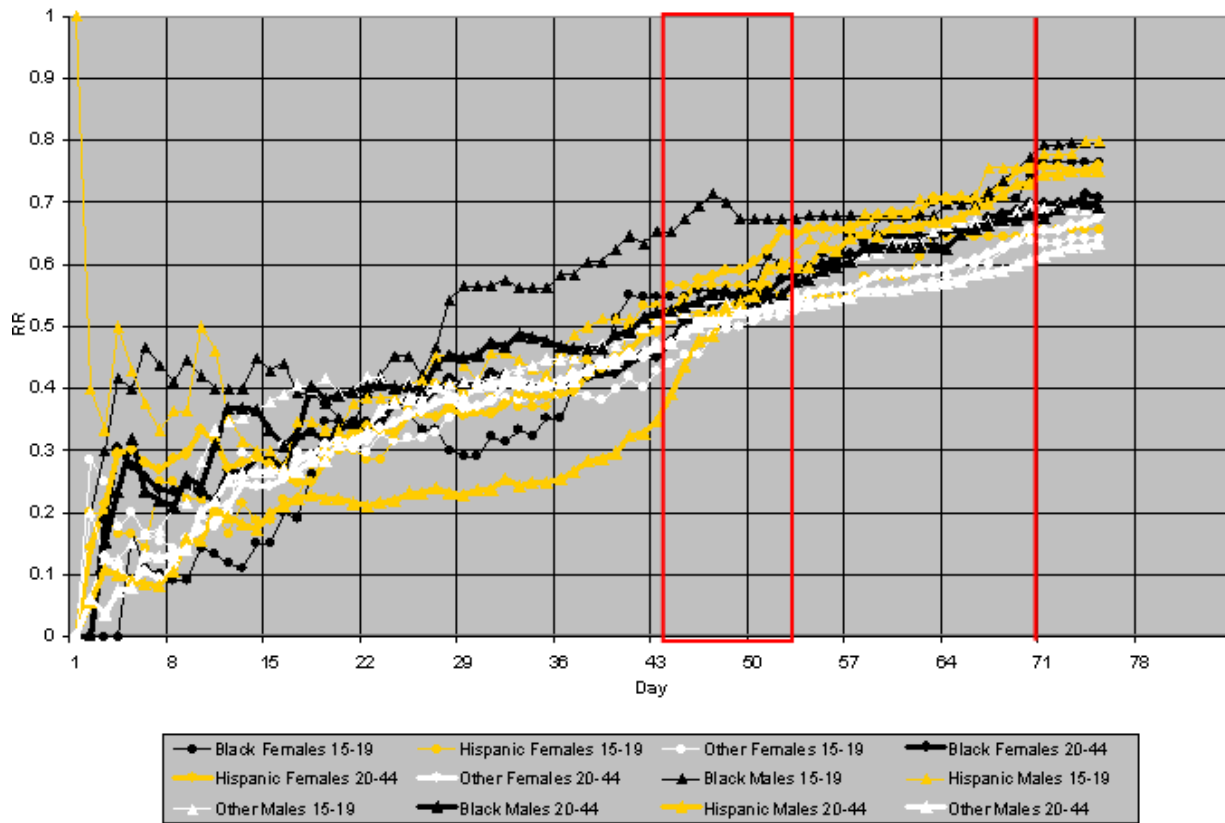
3. Daily Monitoring of Response Rates for Main Interviews across 12 Socio-Demographic Groups

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).

4. Randomized 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 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.

Response Rates by Subgroup



These interventions have been based on a variety of indicators available to use 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.

Starting in the second year of the continuous NSFG, we made the decision that all of these interventions would be randomized to a subset of the cases eligible for the interventions. This provided measurability of the effects of the intervention with traditional statistical analysis. A total of 16 such interventions were conducted during 2006-2010 Continuous NSFG (see Lepkowski et al., under review). Some of these interventions have succeeded in raising response rates for the targeted cases; others have not, as shown in the graph below.

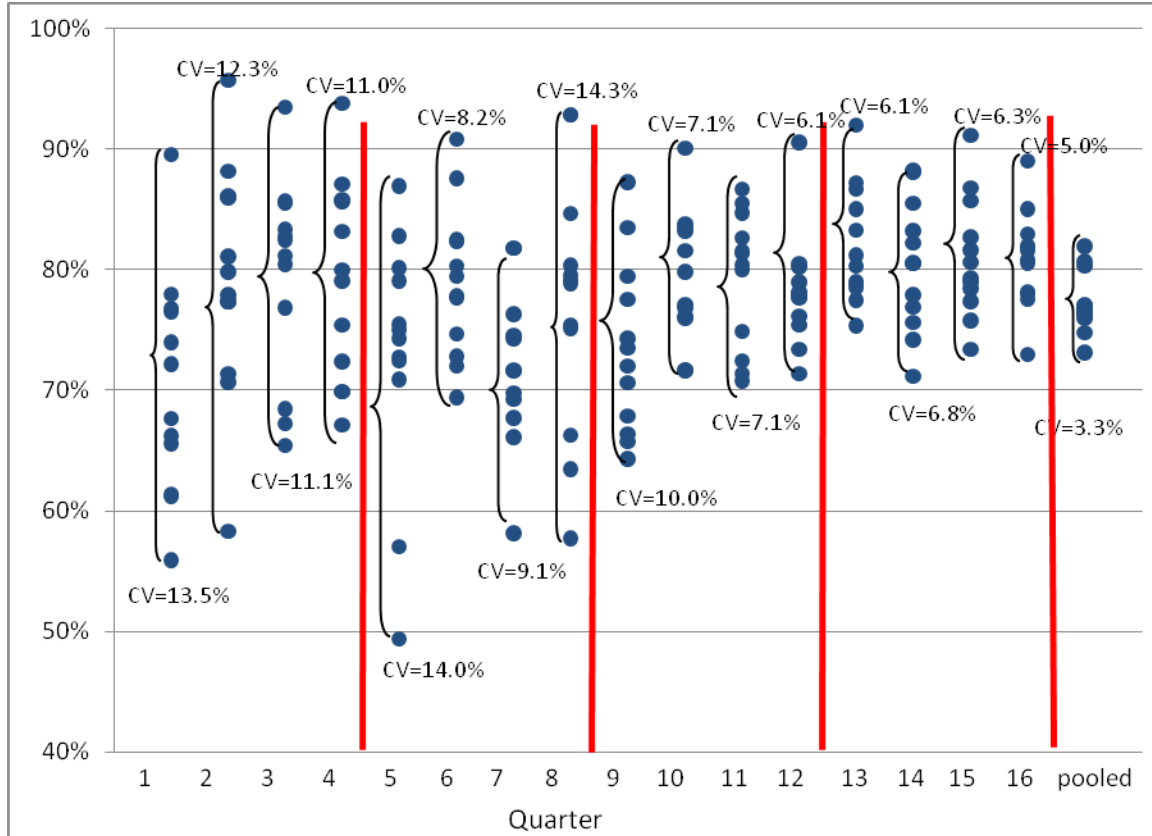
appears to be effective in raising the propensities of the remaining cases, bringing into the respondent pool persons who would have remained nonrespondent without the second phase.

Given the randomized nature of the second phase, we will be able to measure the impact on key NSFG statistics (as we did in Cycle 6; see Axinn, Link, and Groves, 2011). These analyses are currently underway and will inform an evaluation of 2006-2010 Continuous NSFG, and strategies to reduce nonresponse bias in 2011-2015 Continuous NSFG.

6. Comparison of Alternative Postsurvey Adjustments for Nonresponse

We are now examining alternative adjustment models for use with the 2006-2010 Continuous NSFG data set to be released in late 2011. This work has pointed out to us the value in separate uses of auxiliary variables that predict likelihood of response to the survey and those that predict key survey variables. We have estimated correlations between some of the interviewer observations and key NSFG variables and some achieve correlations in the range of .2 to .4. These levels were found to be a minimal requirement for impact on adjustment in a multi-study evaluation of propensity model adjustments (see Kreuter et al., 2010).

One of the positive impacts of the interventions described above was a reduction in the variation of response rates for important subgroups. The graph below shows how this variation decreased over the 16 quarters of the 2006-2010 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. We are currently evaluating the impact of these improved subgroup response rates on key estimates.



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. Given the recent release of the 2006-2010 public use data files (the data were released in October 2011), we are turning to a detailed nonresponse bias analysis of the 2006-2010 data with a view to informing procedures for 2011-2015 Continuous NSFG.

References for Attachment M

- Axinn, W., Link, C., and Groves R.M. 2011. Responsive Survey Design, Demographic Data Collection, and Models of Demographic Behavior. Demography, 48 (3): 1-23.
- Groves R, Benson G, Mosher W, et al. 2005. Plan and Operation of Cycle 6 of the National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 42. Hyattsville, MD: National Center for Health Statistics.
- Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero-, C., Lemay, M., Peytchev, A., Groves, R.M., and Raghunathan, T.E. 2010. Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-Response: Examples from Multiple Surveys. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173 (2): 389-407.
- Lepkowski, J.M., Axinn, W.G., Kirgis, N., West, B.T, Kruger-Ndiaye, S., Wagner, J., and Groves, R.M. 2011. Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection. Under review at the Journal of the Royal Statistical Society.
- 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., et al. 2012a. Results of Fieldwork, Weighting, Imputation and Variance Estimation in the 2006-10 National Survey of Family Growth. Vital and Health Statistics, Series 2, forthcoming. Hyattsville, MD: National Center for Health Statistics.
- Lepkowski, J.M., West, B.T., Wagner, J., Couper, M.P., Kirgis, N., Axinn, W.G., and Mosher, W.D. 2012b. Measuring Nonresponse Bias in the 2006-2010 National Survey of Family Growth. Paper proposed for presentation at the annual meeting of the Population Association of America.
- Wagner, J., Lepkowski, J.M., West, B.T., Couper, M.P., Kirgis, N., Axinn, W.G., and Mosher, W.D. 2012. Examining the Impact of Nonresponse on Estimates from the 2006-2010 Continuous National Survey of Family Growth. Paper proposed for presentation at the annual meeting of the Population Association of America.