



## **ABSTRACT**

Continuous data collection in the National Survey of Family Growth (NSFG) means ongoing sampling of the U.S. age-eligible household population such that interviewing and data production occur daily, with changes in questionnaires and sampling areas undertaken annually. This design seeks to maximize the use of resources and obtain high-quality responses by (1) producing a balanced, continuous work flow, (2) using a small, cross-trained project team, and (3) monitoring data collection in real time to make adjustments to the design. This paper summarizes some of the key design and management strategies used in the successful operation of the NSFG, 2006-2010, highlighting sampling strategies, paradata monitoring techniques, and responsive design initiatives.

## INTRODUCTION

The goal of the continuous National Survey of Family Growth (NSFG) design was to interview a large, nationally representative sample of men and women 15-44 years of age “with greater cost-efficiency and greater control over sample size, data quality, and cost through the use of a more efficient sample design, and extensive use of paradata to make real-time management of interviewer effort possible.” (Groves et. al, 2009).

Continuous data collection in the NSFG context is a design of on-going sampling of the U.S. age-eligible household population so that interviewing and data production occurs daily, with changes in questionnaires and sampling areas annually. This design seeks to maximize the use of resources and obtain high-quality responses by (1) Producing a balanced, continuous work flow, (2) Using a small, cross-trained project team, and (3) Monitoring data collection in real time in order to make adjustments to the design.

The decision to switch from a one-time, large-scale data collection effort to a continuous, small-scale operation was based on a growing body of evidence suggesting that “...despite increasingly costly efforts to make contact with designated households and persuade respondents to participate, response rates have not only continued their decline but also have done so at an increasing rate.” (Singer, 2006)

In addition to increased costs and lower response rates, periodic, large one-time cross-section samples can entail a great amount of risk. These risks take several forms. For example, large studies require a large interviewer force to be assembled and managed. Periodic surveys entail risks of missing measurement of key changes in trends. Because of risks like these, an alternative design was considered, one which would allow for tighter control over costs, provide sensitivity to societal trends, and increase efficiency of data collection. In addition, the design would maximize use of resources and obtain high-quality responses.

To accomplish this, we implemented an over-arching change in the culture of how we conduct data collection—changes to how we staff, changes in the way sample is

allocated, changes to operational protocols, and changes in the way we monitor production and the use of responsive design.

We reduced peak-load staffing burdens through continuous operation with the use of a small, cross-trained project team, and work flow that makes every day identical in terms of interviews obtained, post-processing completed, and data records released.

- Each day a small number of completed interviews are transmitted from the field to Ann Arbor, checked, edited, and placed in a cumulative “raw” data set. New paradata are uploaded; statistical forecast and monitoring models are re-estimated.
- Each week, the interviewer checks on segment listings, screens selected households, and conducts interviews. Ann Arbor staff monitors sample, shifts focus of interviewers to optimize efficiency, makes final decisions on outstanding sample, and checks verification data.
- Each quarter, data collection on one sample ends and a new sample is released.
- Each year, a new set of small sampling areas are rotated into the design and an equivalent set rotated out; new interviewers are hired and questionnaire modules are changed as needed.

Continuous interviewing provides the chance to optimize responsive design over time through forecasting via statistical modeling for responsive design, to reduce risk of the unexpected by a small number of cases being active at any moment, the ability to adjust workloads in sampling areas based on experience in the area, and the use of paradata (process data) monitoring (on hours, costs, outcomes of visits to households) to identify causes of difficulties and adjust field procedures.

This paper summarizes some of the key design and management strategies used for data collection in a continuous environment. Specifically, the paper will highlight the sampling strategies, paradata monitoring techniques, and responsive design initiatives used for the successful operation of the continuous National Survey of Family Growth from 2006 through 2010.

## **BACKGROUND**

NSFG is an ongoing cross-sectional, multistage area probability sample of households. In each sample household, interviewers complete a screening interview by collecting a roster of household members. One person aged 15-44 is selected at random from those who are age-eligible within the household. The interviewer then seeks a 60-80 minute interview from that person. The CAPI questionnaire contains questions on sexual and fertility experiences of the respondent. More sensitive items (e.g., involving risk behaviors for HIV) are administered using ACASI.

Each year of the NSFG consists of four replicate samples, introduced at the beginning of each quarter. The full data collection period lasts 48 weeks, with four weeks stoppage for end-of-year holidays and training of new interviewers. The new interviewers are introduced as part of a rotation of the sample primary sampling units (PSUs) across non-self-representing areas each year. At any one point the sample consists of 25 nonself-representing areas and 8 self-representing areas, with about 38 interviewers working. Additional details about the design and operations of the Continuous NSFG, including detailed descriptions of paradata collected, can be found elsewhere (Groves et al., 2009).

## **METHODS: SAMPLING STRATEGIES**

The sampling strategies used for the continuous management of NSFG included a field-guided design based on an employment model for continuous interviewing and the use of a double sample phase.

### **Employment Model for Continuous Interviewing**

The continuous NSFG utilized an interviewer employment model in which interviewers were required to work 30 hours per week. Based on the fixed number of hours (interviewers staffed x 30 hours per week), a sample design was constructed around the projected interviewer effort.

Instead of a group of over 200 interviewers, which was the model used for the previous cycle of data collection, the field staff for NSFG consisted of approximately 40 field researchers (interviewers) with direct supervision from two field operations coordinators. This was a small, elite, and dedicated team with special recruitment efforts focused on securing staff who were not only experienced in all facets of the work required for NSFG, but who had proven success in field work and had specifically demonstrated leadership abilities. In addition, applicants who had a history of interest in the field of social sciences, either through education, past work experience, or through volunteer work were given special consideration.

Selecting the right staff was critical for this design, given that in most primary sample units only one interviewer was staffed in order to keep the design small and controlled. Therefore, the risk of attrition had to be minimized in order to avoid having unstaffed areas.

Training occurred each year with a new group of recruited staff as the sample areas changed. The training was approximately one week and was designed to be hands-on with the systems needed to complete the work. Background lecture material was moved to DVD format and was completed at home before coming to in-person training. There was consistency of training across the years.

Supervising the two field operations coordinators was a field production manager who was the liaison with the management staff at the central office. The management staff was small, cross-functional, and extremely close to the field work.

When new interviewers joined the project, they were asked to complete a “Field Researcher Questionnaire” in order to collect data on the characteristics of those collecting the data. The changes in focus of recruitment resulted in a staff of interviewers who were significantly more likely to have interviewing experience, higher education, and were more likely to say they like approaching the household (based on an interviewer characteristic questionnaire completed by the interviewer).

### Characteristics of Cycle 6 vs. Cycle 7 Interviewers

Measure	Cycle 6	Cycle 7	Percent Change
Interviewing Experience <i>percent that had past experience</i>	51.63	73.33	21.70**
Education Level <i>percent some college/college grad</i>	59.75	98.33	38.58***
Ability to Figure out Respondent Objectives <i>percent answered strongly agree/agree</i>	86.99	91.67	4.68
Like Approaching Household <i>percent answered 8-10 where 10 is best</i>	38.62	56.67	18.05*
Importance/Relevance of Work <i>percent answered 8-10 where 10 is best</i>	89.43	93.33	3.9

\*.01 < p ≤ .05

\*\* .001 < p ≤ .01

\*\*\* p ≤ .001

### Field-Guided Design

In order to build a sample design around the expected effort of field work, we used Cycle 6 information at a Primary Sampling Unit level to set expectations for the hours per interview needed. For example, if a realized hours per interview in a certain part of the country was 8.6 in the previous cycle, we used this as the basis of the effort required for Cycle 7, which was in the same areas as the previous cycle. For an interviewer working 30 hours per week for 12 weeks in an area with an estimated hours per interview of 8.6, we would expect approximately 42 completed interviews. The number of housing units to be selected in the area was then determined factoring in further details such as occupancy rates, eligibility rates, and response rate in order to get to a final sample size that would both provide adequate work for the interviewer while meeting the objectives of the study.

Once an initial allocation of housing units was drafted for each area, a series of review and adjustments were made each quarter in order for the field management staff to contribute additional design considerations at the sample area level. For example,

adjustments up or down in the allocation of the lines might be made based on specific information about the interviewer working in the area and the type of area/conditions in the area.

### Use of a Double-Sample Phase

Data collection was completed in a series of four quarters each year, each quarter lasting 12 weeks. The first ten weeks of the quarter was “phase 1” of the study and the last two weeks of the quarter was “phase 2.”

In phase 1, normal protocols applied. Selected housing units were screened for eligibility and interviewed when eligible. In week 10, the sample was reviewed and a sub-selection of the remaining active lines were designated for phase 2, typically 30% of the remaining lines were selected for continuation in the second phase.

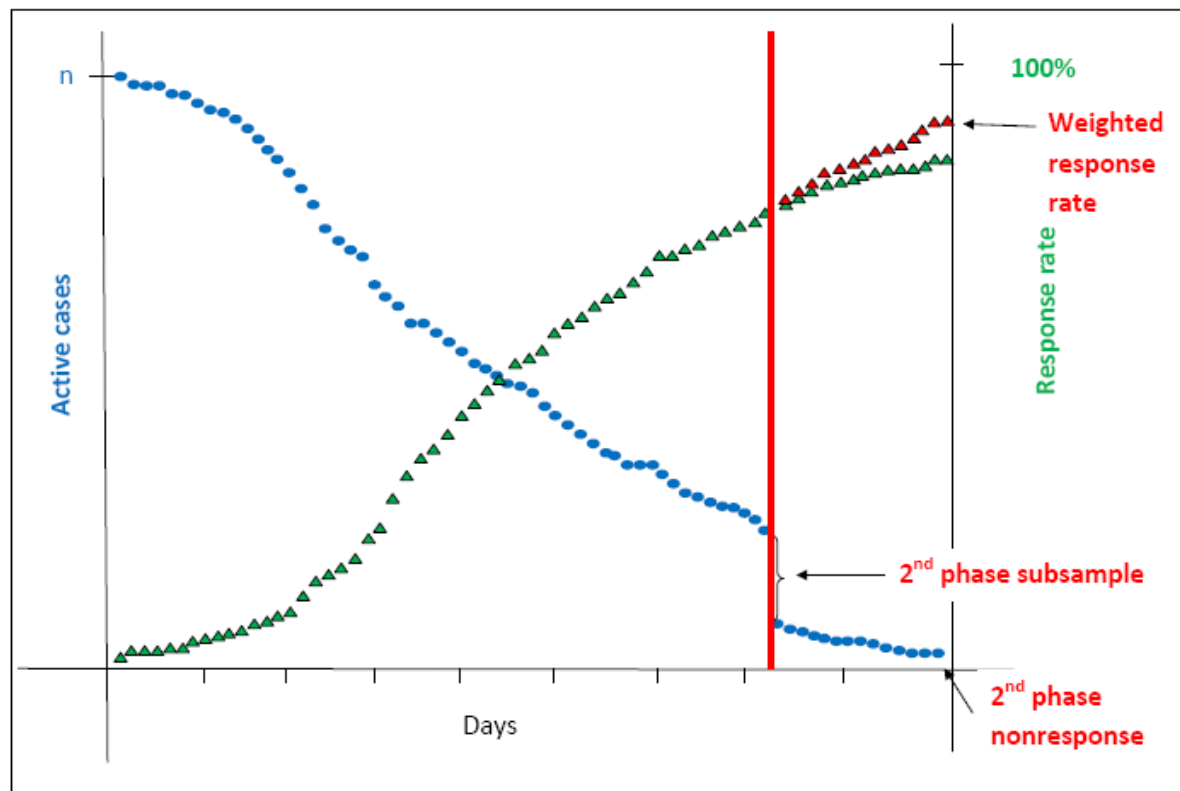
In order to select the double sample, active cases were stratified by likelihood of response, by type of line (screener line or main line), and by selection weights. Those with a higher likelihood to respond were oversampled and those with a lower likelihood were under sampled. Lines with a larger sampling weight had a larger probability of selection, which reduced variation in sampling weights in final weighted sample.

In phase 2, only the sub-selected lines were worked and all other lines were coded out of the sample and a change in recruitment protocol was implemented; Adults in the double sample were given an increased token of appreciation (from \$40 to \$80) and the use of proxy reporters for the completion of screeners was allowed. In addition to the change in recruitment protocol, there was also a change in interviewer behavior as well, with all of the interviewer effort applied to very few lines in order to maximize response rates with the remaining cases.

The chart below provides an illustration of the drop off in the number of remaining lines in the second phase subsample and the gain in the weighted response rate as a result of this sampling strategy. In the last two weeks of the 12-week quarter, the phase 2 sample lines received an extra token of appreciation offered to adult respondents and the contact protocols are revised.



Illustration of Second Phase Sample and Response Rate



There are advantages and disadvantages to the use of a double sample phase in data collection. The advantages include reduction in exponential cost inflation at end of survey, control over high effort cases, control over costs, and change of recruitment protocol – more effective design. The disadvantages include weights to compensate for second phase sample and potentially higher sampling variance in estimates.

### METHODS: PARADATA MONITORING TECHNIQUES

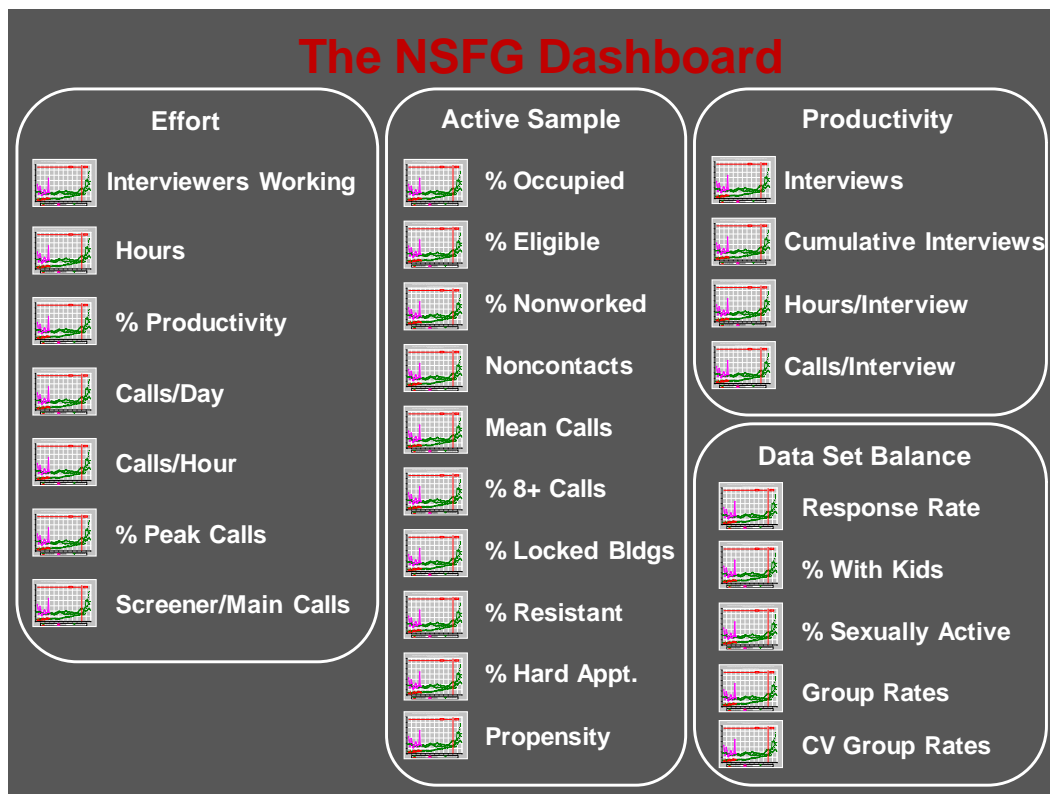
Paradata comes from several levels: Interviewer (for example, hours worked), interviewer observations of the segment, housing unit (for example, locked buildings), observations from each visit to the household (time of day, day of week, outcome of call), observations on each contact with someone in the household, observations on each interviewer day (travel time to segment, time spent attempting screeners and main interviews)

For NSFG, we created a “dashboard” of key paradata indicators for daily monitoring of the study outcomes, using a production model that motivates paradata considerations.

The product of the production process is costs and completed interviews. Both screener and main interviews are valuable, but the data records are based on main interviews, making screener interviews a necessary but insufficient intermediate product. This production model specifies that the success of any field effort to produce an interview at some target cost is a function of the quality of the current active sample cases (i.e., sample addresses whose status is not yet finalized). At the beginning of the quarterly NSFG data collection period, all cases are in an “untouched” status, not having ever received a contact attempt. The average cost of producing an interview is lowest at this point. On the last day of the data collection period, the few remaining active cases have typically received many calls, and many calls across many cases are required to produce one interview. The likelihood of producing an interview and the cost of producing an interview with each act of interviewers is thus a function of the “quality of materials.”

In sum, the production model says that the cost and likelihood of an interview is a function of the field effort applied and the current state of the active sample case base.

**The NSFG Dashboard**



Review of production measures on effort, quality of the active sample, productivity, and data set balance occur daily. At different points in each quarter, some production measures are focused on more than others.

Early Quarter (Weeks 1 through 3) Indicators: The focus in this period is on management of the effort, not the product. Careful management of the input at this phase in the quarter will ensure that the product will result as work progresses in the quarter. Screening households early in the quarter is pivotal in order to gain an assessment of the sample. Therefore carefully tracking the number of interviewer hours being worked and when interviewers are working is crucial. Also important during this period is the calling levels, calling in peak calling windows, the ratio of screener to main calls, the number of hard appointments, and ensuring that all lines have been worked (carefully tracking the number of non-worked screener lines).

Middle Quarter (Weeks 4 through 6) Indicators: At this point in the quarter, effort continues to be monitored closely, but we also begin to focus on the quality of the active sample. Eligibility rate of the households, the number of lines in locked buildings, level of resistance on contacted lines, calling levels, and sample line propensities.

Late Quarter (Weeks 7 through 9) Indicators: Productivity and data set balance become the focus at this point in the quarterly data collection effort by evaluating the output for the closely-monitored input in the early phase of data collection. Key productivity indicators are interview counts, hours per interview, and calls per interview. Data set balance indicators include response rates, and sub-group response rates in an effort to reduce the risk of nonresponse bias. In addition, the evaluation of respondents and non-respondents on auxiliary indicators, such as households with children and respondents in a sexual relationship with an opposite sex partner assist with data set balance. "Collecting auxiliary variables on respondents and non-respondents to guide attempts to balance response rates across key subgroups is wise." (Groves, 2006) .

## **METHODS: RESPONSIVE DESIGN INTERVENTIONS**

Groves and Heeringa (2006) define responsive design: “The ability to monitor continually the streams of process data and survey data creates the opportunity to alter the design during the course of data collection to improve survey cost efficiency and to achieve more precise, less biased estimates. We label such surveys as ‘responsive designs’.”

Groves and Heeringa further outline the characteristics of responsive designs:

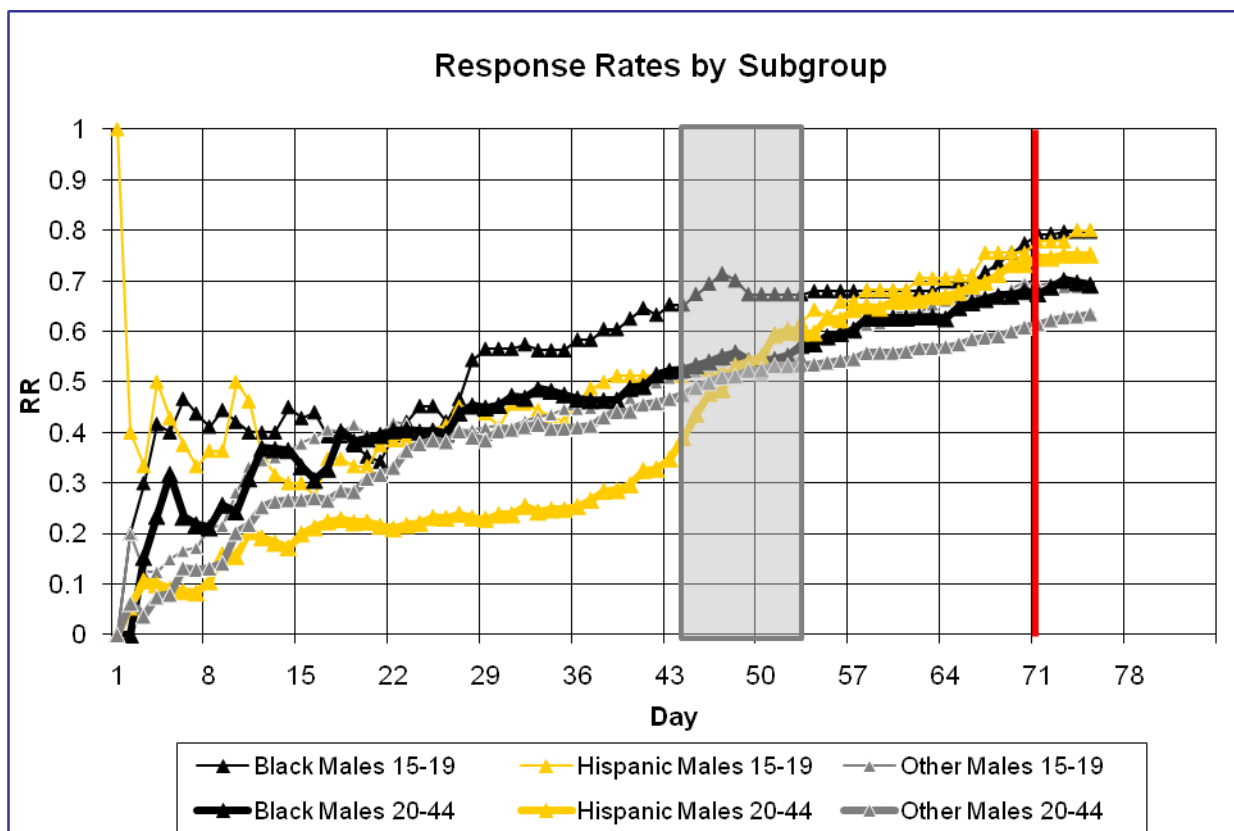
- a. pre-identify a set of design features potentially affecting costs and errors of survey estimates.
- b. identify a set of indicators of the cost and error properties of those features and monitor those indicators in initial phases of data collection.
- c. alter the features of the survey in subsequent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

Responsive design requires active management of the field effort through the use of paradata and paradata monitoring. With responsive design, interviewer effort is used with maximum effectiveness and study procedures can be changed in response to changing field conditions. The effects of these changes ("interventions") can be documented by paradata.

The NSFG oversamples certain populations, such as Hispanics, in order to have sufficient cases for estimation. One type of responsive design intervention on NSFG involves targeting a specific subgroup of active sample cases found to have poor or lagging values on key process indicators. For example, the figure shows that between days 1 and 43 of Quarter 14, older (age 20-44) male Hispanics were lagging in terms of response rates. In response to this observed trend (monitored daily from the NSFG dashboard), NSFG managers implemented an intervention designed to have interviewers

target this specific subgroup of sample lines. This intervention began on day 44 of the quarter, with targeted older male Hispanic lines flagged in the sample management system.

The chart shows the relatively sharp increase in response rates within this subgroup over the next week. This intervention had the beneficial effect of decreasing the coefficient of variation (CV) in the response rates among these six subgroups, which is a very important process indicator monitored by NSFG managers to assess balance in the data set. After this intervention was completed, the coefficient of variation of the response rates remained stable for the remainder of the quarter.

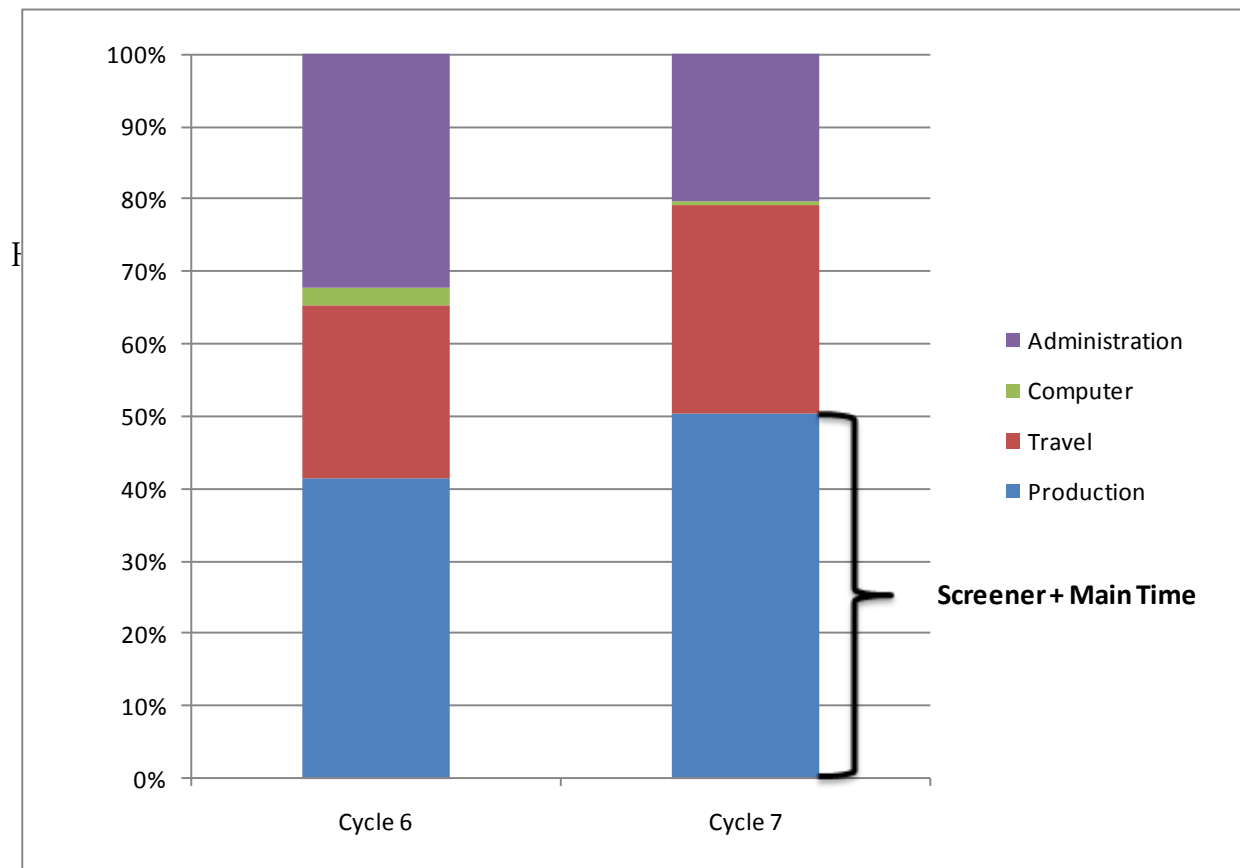


## RESULTS

The change from a one-time, large-scale data collection effort to a smaller, continuous design required changes in the employment model for interviewers, sample design, paradata monitoring techniques, and the use of responsive design. These changes resulted in the successful implementation of NSFG continuous data collection from 2006 to 2010, with an increase in field efficiency contributing to higher than anticipated interview yield.

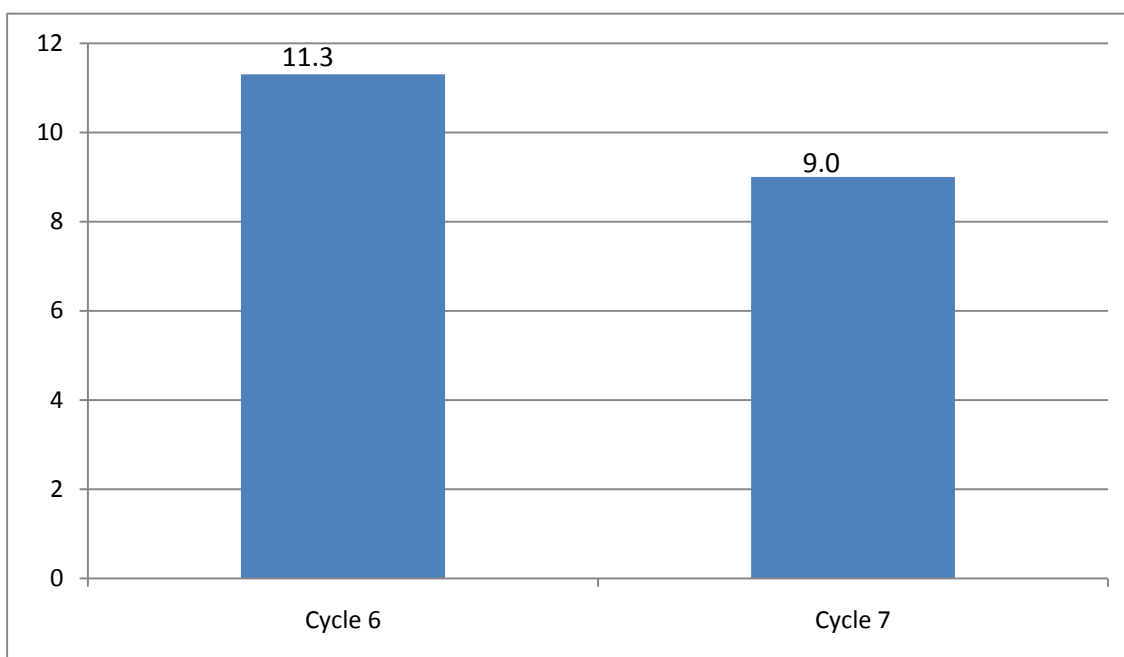
An example of the increase in field efficiency between Cycle 6 and Cycle 7 is the distribution of interviewer hours across various tasks. In Cycle 6, about 40% of interviewer hours were spent on screening and interviewing. The remainder of the hours were spent on non-production activities such as administration time, travel, computer problems. In Cycle 7, about 55% time spent in screening and interviewing.

**Cycle 6 versus Continuous NSFG Percentage Distribution of Interviewer Time**



Another example of the gains in efficiency is demonstrated by the hours per interview. Hours per interview is calculated by the sum of the interviewer hours divided by the total number of interviews collected. The hours encompass all of the tasks necessary to complete the work—administrative, travel, screening households, and conducting interviews. For Cycle 6, each completed main interview required 11.3 hours of work. In Cycle 7 with the continuous design, the achieved hours per interview is 9.0. It is worth noting that other operational refinements to the data collection protocols were put into place for continuous interviewing which also contributed to the improvement in efficiency and cost savings. For example, tablet computers were used for electronic listing of sample segments and for obtaining signatures on consent forms. Using tablet computers for these tasks eliminated the need for paper processing of sample lists and consent forms.

#### Cycle 6 vs Continuous NSFG Interviewer Hours to Complete a Main Interview (Cumulative)



The increased efficiency realized through the continuous interviewing design resulted in a 79% increase in interview yield between Cycles 6 and 7. In Cycle 6, ISR collected 12,571 main interviews in a twelve-month data collection period from March 2002 through February 2003 with a response rate was 78%. In Cycle 7, we will collect approximately 22,500 main interviews in a data collection effort spanning four years, from July 2006 through June 2010, with a response rate of 76%.

## REFERENCES

- Groves, Robert M., Benson, Grant, Mosher, William D., et al., "Plan and Operation of Cycle 6 of the National Survey of Family Growth." National Center for Health Statistics, Vital Health Statistics, Series 1, Number 42. 2005.
- Groves, Robert M., Mosher, William D., Lepkowski, James M., and Kirgis, Nicole G., "Planning and Development of the Continuous National Survey of Family Growth." National Center for Health Statistics, Vital Health Statistics , Series 1, Number 48. 2009.
- Groves, Robert M., and Heeringa, Steven, G., "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Nonresponse and Costs," *Journal of the Royal Statistical Society*, 2006.
- Groves, Robert M., "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly*, 2006.
- Singer, Eleanor, "Nonresponse Bias in Household Surveys," *Public Opinion Quarterly*, 2006.