

Note: This is same attachment as used in the 2015-2018 and 2018-2021 clearance request.

## **Incentive Experiments in the National Survey of Family Growth (NSFG)**

### **Summary**

The National Survey of Family Growth (NSFG) is the national fertility survey of the United States, gathering nationally-representative data on the factors that affect birth and pregnancy rates in the US—including sexual activity, contraception, marriage, and infertility. The survey has a history of offering incentive payments to respondents, and of doing experiments to test the effectiveness of incentives. This attachment shows that the experiments done to date clearly establish that incentives reduce costs in the NSFG, that they increase response rates, and they increase the representativeness of the NSFG sample.

The most recent experiments in 2002-3 and 2006-7 suggested that without an increased incentive for a small proportion of respondents (6-8%), groups such as childless and college-educated women, and Hispanic men are not as well represented in the standard Phase I sample as they are when \$80 is offered to a sub-sample of non-respondents. Bringing these groups into the sample improves the representativeness of the sample, and raises the response rate, while avoiding the high costs of repeated visits to non-responding households. This appears to justify the use of the \$80 amount for a small sub-set of the sample. Field conditions have not changed materially since that time, so we seek to continue the incentive structure used in 2007-2010 for this 3-year period, or until field conditions necessitate a change. If that occurs, we will propose a new experiment to OMB.

### **Capsule History of Incentive Payments in the NSFG**

NSFG has a history of providing incentive payments to respondents, and of testing alternative levels of payment. (All NSFG interviewer materials and consent forms refer to the incentive as a “token of appreciation.”) Incentives in the NSFG are in the form of cash payments at the time the interview begins. Four major experiments with incentives have occurred in the NSFG. We will describe the first three of these briefly, and then describe the most recent one in detail:

- 1) **1993 (Cycle 5) Pretest:** In a field experiment in the 1993 pretest for NSFG Cycle 5, a \$20 cash incentive was found to produce a significantly higher response rate (67.4%) than when no payment was offered (58.9%). For women who were offered \$20, response rates were higher, and field costs per case were lower than for women who received no incentive.
- 2) **2001 (Cycle 6) Pretest:** In a field experiment in the 2001 Pretest for Cycle 6, a \$20 payment was contrasted with a \$40 payment. The response rate for those offered \$20 was 62%, and for those offered \$40, it was 72%. There was variation in the differences across demographic groups as well. Women offered the \$20 incentive had a response rate of 62%, while women offered \$40 incentive had a response rate of 81%. Those receiving the higher amount were also less likely to express objections or reluctance to the interview than those receiving \$20.
- 3) **Cycle 6 Main Study:** In the 2002-2003 Cycle 6 Main Study, a \$40 incentive was used, but response rates were still lagging in key groups after seven months of interviewing. NSFG staff requested and received from OMB permission to use an \$80 incentive in a half-sample of the cases remaining in the final four weeks of data collection during February, 2003. The \$80 incentive raised the weighted response rate from 64% to 79%. The sample in the last 4 weeks had a higher proportion of married women, Hispanic men and women, and full-time workers of both sexes.

These experiences showed cost-effective increases in response rates and representativeness when incentives were offered to potential respondents.

### **Experiment 4: 2006-2007**

The fieldwork for the 2006-2010 Continuous NSFG comprised two stages: obtaining a simple household roster to see if someone 15-44 years of age lives in the household (the “screener”), and obtaining a main study interview. Each data collection year is divided into four 12-week quarters. For 10 weeks, interviewers offered a \$40 incentive during attempts to obtain

interviews. The response rate by the end of the 10 weeks in Year 1 of interviewing averaged about 60 percent, judged too low by the National Center for Health Statistics and our co-sponsoring agencies. By the end of week 10, interviewers had visited households an average of eight times in person to obtain a screener and an interview. Given the cost of these visits in interviewer time and expenses, something had to be done to improve the odds of success.

In week 11 of each quarterly data collection period (the beginning of “Phase 2”), study staff drew a sub-sample of the remaining unfinished cases. The remaining cases were generally of two types:

- (a) non-contacted cases (such as completing a screener with Mrs. Jones; Mr. Jones is selected as the respondent, and the interviewer is unable to find him at home for an interview);  
and
- (b) “soft refusals” (such as “I can’t do it this week because I’m sick,” or “we’re having dinner now, come back later”).

Initially, we sought to use the same procedure that was approved by the Office of Management and Budget and the NCHS IRB, and used successfully in February, 2003 in Cycle 6 of the NSFG: Phase 2 cases were offered an \$80 incentive (instead of \$40), in two payments.

- The first \$40 was prepaid and delivered in a Fedex letter.
- The remaining \$40 was paid at the time the respondent sat down to do the interview.

In Cycle 6 (2002-3), 724 respondents out of 12,571, or 6%, received the higher incentive.

For the continuous NSFG starting in 2006, NCHS and University of Michigan staff sought to apply this same token payment system. It was anticipated that a higher incentive during Phase 2 would in the end save money and improve the representativeness of the sample. So we proposed to use a \$40 incentive during Phase 1 of the Continuous NSFG, and an \$80 payment in the Phase 2 non-response follow-up.

However, NCHS management suggested that we had not demonstrated that it was necessary to raise the incentive from \$40 to 80 in weeks 11 and 12. They suggested that NSFG conduct an

experiment to see if we could increase it from \$40 to \$50 (instead of from \$40 to \$80) and obtain a similar improvement in response rates, while reducing the overall cost of data collection.

A randomized experiment was thus conducted during from September of 2006 to June of 2007 (quarters 2, 3, and 4) to determine whether a \$50 incentive would prove as effective as an \$80 incentive in increasing response rates and yielding a sample with better balance across key subgroups.

### **Design of the Experiment**

The basic experimental design operated within the Continuous NSFG 12-week quarter. During Phase 1 (weeks 1-10 of each quarter), potential respondents were offered a \$40 incentive to complete an interview. During Phase 2 (weeks 11 and 12 of each quarter), a “double sample” of approximately one-third of the remaining (non-responding) cases was selected. Some sample cases selected into Phase 2 were still at the screener stage, and others were at the main interview stage. The Phase 2 sample in quarters 2, 3, and 4 was then randomly divided into two groups:

- Group 1 received \$10 prepaid in addition to the standard \$40 (a total of \$50 for the main interview);
- Group 2 received \$40 prepaid in addition to the standard \$40 at completion of interview (a total of \$80 for the main interview).

If a household in either group had not completed a screener at the end of Phase 1, they were offered a \$5 prepaid token to complete the screener in Phase 2. These two groups are designated as the \$5/\$10/\$40 and the \$5/\$40/\$40 experimental conditions respectively. For brevity, we will refer to these as the \$50 and \$80 groups respectively (only a small subset got the additional \$5 incentive for completing the screener.)

Cases selected for Phase 2 were sent a final letter via express mail with the prepaid token enclosed. The letter stated that the enclosed token was for the respondent to keep, in appreciation for their help.

**Table 1** below presents the pooled results across quarters. The table shows counts of cases and response rates separately for household screener and main interview cases to be able to evaluate the potential impact of the incentives in each group.

**Table 1. Unweighted Case Counts for Phase 2 Incentive Experiment Outcomes, with Response Rates and Standard Errors, September 2006-June 2007 NSFG**

<b>Screener Interview Cases in Phase 2</b>							
Experiment	Sample size	Completed Screener Interviews	Refusal	Non-Interview	Non-Sample	Response Rate	Std Error
<b>\$50</b>	208	130	43	20	15	<b>67%</b>	3.4%
<b>\$80</b>	207	152	32	14	9	<b>77%</b>	3.0%
<b>Main Interview Cases in Phase 2</b>							
Experiment	Sample size	Completed Main Interviews	Refusal	Non-Interview	Non-Sample	Response Rate	Std Error
<b>\$50</b>	192	100	48	44	0	<b>52%</b>	3.6%
<b>\$80</b>	215	137	29	49	0	<b>64%</b>	3.3%

Minors 15-17 were not included in the experiment. Their token of appreciation for the interview was never more than \$40. Randomized assignment of Phase 2 cases to treatment groups was made on the segment level (i.e., all cases in a segment were assigned to the same treatment group). Therefore, the simple random sample standard errors are likely to underestimate the true standard errors.

The response rates for screener cases in Phase 2 that were offered the \$40 pre-paid main interview incentive were 10 percentage points higher than for the screener cases offered the \$10 pre-paid main (77% vs. 67%). The response rates for main interview cases in Phase 2 were 12 percentage points higher in the \$80 group than the \$50 group (64% vs. 52%).

A variety of hypotheses were tested concerning sample composition differences between the groups. Table 2 shows results of various hypothesis tests applied to the experimental data to examine differences in the composition of the two responding groups. The sample sizes in the experimental categories are small because only a one-third sub-sample of the remaining cases were selected for Phase 2, and that sub-sample was randomly split into two payment plans.

**Table 2. Comparison of sample characteristics between the \$50 and \$80 experimental groups**

**in the NSFG in Sept 2006-June 2007 (Quarters 2, 3, and 4)**

Characteristic	Phase 1	Phase 2	
		\$5/\$10/\$40 (\$50 total)	\$5/\$40/\$40 (\$80 total)
<b>Female</b>			
<b>Sample size</b>	<b>1,896</b>	<b>51</b>	<b>68</b>
College degree, or more	34%	48%	51%**
Ever had an abortion	6%	3%	1%**
Ever had a live birth	59%	68%	40%**
Ever had sex with a female	13%	16%	4%**
Income \$75,000 per year or more	17%	40%**	25%
Living in a multi-unit structure	38%	24%*	24%**
<b>Male</b>			
<b>Sample size</b>	<b>1,432</b>	<b>47</b>	<b>70</b>
Hispanic	20%	24%	37%**
College degree, or more	28%	43%*	36%
Ever fathered a birth	43%	48%	36%
Ever had sex with a male	7%	5%	1%**
Income \$75,000 per year or more	25%	30%	42%**
Living in a multi-unit structure	37%	42%	26%*
Physical impediments to housing unit	12%	10%	20%

\*Two-sided hypothesis test comparing Phase 1 vs. Phase 2 experimental group statistically significant at the  $\alpha = 0.10$  level.

\*\*Two-sided hypothesis test comparing Phase 1 and Phase 2 experimental group statistically significant at the  $\alpha = 0.05$  level.

The sample sizes in the two Phase 2 experimental groups are small (for women,  $n=51$  in \$50 group and  $n=68$  in the \$80 group), but the sample size in the phase 1 group is large ( $n=1,896$ ). For men, the sample sizes are similar: 47 in the \$50 group and 70 in the \$80 group, but 1,432 in the \$40 Phase 1 group.

Despite the small sample sizes in the Phase 2 groups, 9 of the 13 differences between Phase 1 (\$40) and the \$80 group were significant, using 2-tailed t-tests, while only 3 of the 13 comparisons between the \$40 group and the \$50 group were significant. These results suggest that the \$80 incentive was recruiting different people into the sample, but the \$50 incentive was much less effective in that respect. Given due caution about the sample sizes, the patterns are clear, and quite similar to those found in the final phase of Cycle 6:

For women:

- A significantly larger proportion of women offered \$80 group had a college degree (51%) than the group of women offered \$40 (34% with college degrees).
- A significantly smaller proportion of women offered \$80 had one or more live births (40%) than women offered \$40 (59%). In other words, 60% of the women offered \$80 were childless, compared with just 41% of the women offered \$40. Since the principal outcome variable of the NSFG is fertility and birth rates, this is a critical finding: the higher incentive was more effective at including childless women in the survey. Not using the \$80 amount would bias the sample toward those with children.

For men:

- A significantly higher proportion of men offered \$80 were Hispanic (37%) than men offered \$40 (only 20% of men offered \$40 were Hispanic). Given strong policy and program interest in the Hispanic population, this is also a key finding.
- Men with incomes of \$75,000 or more were 42% of the \$80 group, compared with just 25% of those offered \$40. Higher-income men apparently responded to the \$80 incentive but were less responsive to the \$40 incentive.

For both men and women:

- Those in the \$80 group were less likely to live in multi-unit structures than the \$40 phase 1 group. Among women, 38% of those in the \$40 group lived in multi-unit structures, compared with just 24% of women in the \$80 group.
- Among men, 37% of men in the \$40 group lived in multi-unit structures, compared with 26% of men in the \$80 group. In other words, those in the \$80 group were more likely to live in single-family homes than those in Phase 1 of the sample.

- Also for both men and women, those in the \$80 group were less likely to have had same-sex sexual contact than those in the phase 1 \$40 group (4% of women in the \$80 group had had sex with another woman, vs. 13% of women in the phase 1 \$40 sample).
- For men, 1% of those in the \$80 group had had sex with another man, compared with 7% of men in the \$40 phase 1 group. Given the strong interests in these behaviors for public health reasons, these are also critical findings.

### **Conclusions from the 2006-7 Experiment**

Despite relatively small samples in the two experimental groups, consistent results were obtained across three consecutive quarters: the \$80 incentive raised response rates and recruited different people into the sample than the Phase 1 effort alone (\$40 incentive) or the \$50 incentive. Further, the results are broadly consistent with findings from Cycle 6 (2002 and 2003). The results suggest that busy, college-educated, childless women, and high-income men and Hispanic men, are not as well represented in the standard Phase I sample as they are in the \$80 follow-up sample. It takes the \$80 amount to bring more of these people into the sample. Bringing them in improves the representativeness of the sample, and raises the response rate. This appears to justify the use of the \$80 amount for a small sub-set of the sample.

Given the costs of running the experiment, NCHS and study staff asked permission to end the experiment after three quarters of data collection, for the following reasons:

1. The significant operational costs of the experiment. The experiment increased fieldwork costs because it required two different mailings, randomized assignment of segments, checking that interviewers applied the correct protocol, and analysis of the experimental data. It also increased the complexity of the task for interviewers.



2. To the extent that one of the experimental groups achieved lower response rates and less representative sample composition, the overall response rate and quality of the NSFG sample was being reduced as long as the experiment continued.
3. While it was possible that the response rate advantage of the \$80 incentive treatment would dissipate over time, it was very unlikely given that previous experiments in the NSFG consistently found that higher incentives increased response rates in a cost-effective way.
4. Staff estimated that it would take at least one more year of continuing the experiment to obtain statistically significant differences in response rates and more adequate sample sizes for the sample composition comparisons.

Approval to end the experiment was granted by the National Center for Health Statistics (NCHS) IRB, the NCHS "Research Ethics Review Board," on August 29, 2007 (Amendment 11, NCHS Protocol Number 2006-01) and by OMB in November, 2007. The \$80 token in Phase 2 was adopted in all subsequent quarters of the 2006-2010 Continuous NSFG.

Given that the 2011-2015 Continuous NSFG design is essentially a continuation of the 2006-2010 Continuous NSFG, and we see no changes in the survey climate or essential survey conditions to suggest that the incentives should perform differently going forward, we believe the time and effort needed to mount another incentive experiment is hard to justify, and hence request that OMB accepts the results of the experiments described above as sufficient to justify continuing with the current incentive plan.