

Attachment C: Nov 19, 2007 Memo to OMB

November 19, 2007

To: Margo Schwab
Office of Management and Budget

From: William Mosher, National Center for Health Statistics, CDC/DHHS
Robert Groves, University of Michigan/ISR

Through: Mary Moien, Clearance Officer, NCHS

Subject: **OMB Number 0920-0314**
Results of Incentive Experiment in Year 1 of Cycle 7 NSFG

This is a response to the questions asked by OMB in response to our memoranda of September 6, and October 12, 2007, and our telephone conversation on November 7. We are asking permission to end the experiment with incentives. Permission to end the experiment was granted by the CDC/NCHS IRB on August 29, 2007 (Amendment 11, Protocol 2006-01).

AN OVERVIEW OF NSFG PROCEDURES

The National Survey of Family Growth (NSFG) is the national fertility survey of the United States. That means that it gathers nationally-representative data on the factors that affect birth and pregnancy rates in the US—including sexual activity, contraception, marriage, and infertility. It also gathers data on adoption, men’s roles in families, and behavior related to HIV and Sexually transmitted disease infection. The NSFG is done in person because it is complex and sensitive; it is successful because marriage, children, family, and relationships are important to most people.

One key feature of the continuous design is the cost-efficiency that stems from drastically reducing the number of PSU’s worked in any given year. The sample for Cycle 7 (continuous interviewing) of the NSFG is a sample of 108 Primary Sampling Units (PSU’s). But to increase efficiency, these 108 are divided into 4 annual sub-samples. Each annual sample is a nationally representative sample of 33 PSU’s: (1) Eight PSU’s (the “Big 8”) that are in the sample every year (New York, LA, Chicago, etc), and (2) 25 others. Thus, the sample is about 33 PSU’s in year 1, with the “Big 8” and 25 others, the “Big 8” and 25 new PSU’s in Year 2, and so on. This smaller number of PSU’s each year means lower costs for recruiting, staffing, and management.

The fieldwork comprises two stages: obtaining a simple household roster to see if there is someone 15-44 years of age in the household (“screener”), and obtaining a main study interview. Each year is divided into four 12-week quarters. For 10 weeks, we offer a \$40 incentive and make many attempts to contact households and obtain interviews. Our response rate by the end of the 10 weeks in Year 1 of interviewing has been averaging about 60 percent (which would be judged too low by our co-sponsoring agencies, and by us.) By

the end of week 10, we have visited the average household eight (8) times in person to attempt to get a screener and an interview. Given the cost of these visits in interviewer time and expenses, something has to be done to improve our odds of success.

In week 11 (the beginning of “Phase 2”), we take steps to control costs and obtain a more representative sample. We take a sub-sample of the remaining unfinished cases, excluding final refusals and teenagers 15-17 years of age. Final refusals are not re-contacted. We have agreed with the NCHS IRB that we will not re-contact final (firm) refusals or teenagers 15-17 who have not completed an interview at the end of the first 10 weeks.

The remaining cases, aged 18-44, are 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, “ or “we’re having dinner now, come back later”).

We request to use the same procedure that was approved by OMB and the NCHS IRB and used successfully in February 2003 in Cycle 6 of the NSFG: the selected cases are offered a higher incentive (\$80 instead of \$40), with \$40 of it being prepaid and delivered in a Fedex letter. In Cycle 6, 724 respondents out of 12,571, or 6%, received the higher incentive. (The experiment in Cycle 7, which were are seeking permission to end, split this 6% between 2 groups, with about 3% receiving \$40 in the Fedex letter and 3% receiving \$10.)

This 2-week effort typically raises the weighted response rate from 60% to about 75% in just 2 weeks, while controlling our costs and bringing people with different characteristics into the sample (more details below). Less than 10% of the original sample gets an \$80 incentive. These incentives save money and improve the representativeness of our sample (details below).

HISTORY OF TESTING INCENTIVES IN THE NSFG

CYCLE 5 PRETEST--In a field experiment approved by OMB in the Cycle 5 pretest in 1993, a \$20 incentive was found to produce a significantly higher response rate (67.4%) than no incentive (58.9%--a difference of 8.5 percentage points), as well as lower field costs. (Series 1, No. 36, page 9; and Mosher et al, 1994, ASA).

CYCLE 6 PRETEST---In a field experiment in the Cycle 6 Pretest in 2001, a \$20 incentive was compared with a \$40 incentive. The response rate for those offered \$20 was 62%; for those offered \$40, it was 72%. The effect of incentives at this level was larger for women. Women offered \$20 had a response rate of 62%, while women offered \$40 had a response rate of 81%. (Series 1, No. 42, page 13, tables J and K.) Those receiving the higher amount were also less likely to express objections or reluctance to the interview than those receiving \$20. (Ibid, tables L and M.).

CYCLE 6 MAIN STUDY--In the Cycle 6 Main Study, a \$40 incentive was used, but response rates were still lagging in certain groups after 7 months of interviewing. We requested, and received from OMB, permission to use an \$80 incentive in a half-sample of the remaining cases, for 4 weeks at the end of interviewing in Cycle 6 (February, 2003). That \$80 incentive raised our response rate from 64% to 79%. (Series 1, No. 42, pages 40 and 50, e.g.; presentation to OMB, January 9, 2006, slides 18-20). Furthermore, 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 (presentation to OMB, Jan 9, 2006, slides 19-20).

This experience, showing cost-effective increases in response rates and representativeness, led us to propose using the \$40 incentive with an \$80 non-response follow-up in Cycle 7. However, the NCHS Associate Director for Science suggested that we conduct an experiment to see if it was necessary to increase the payment from \$40 to \$80, or whether we could increase it from \$40 to \$50 and get the same results.

So, during the first 10 weeks of each 12-week quarter, we offered all respondents a \$40 incentive. During the last 2 weeks of each quarter, under this experiment, we took a half-sample of the remaining respondents; half of these were offered an extra \$10 (for a total of \$50), and half were offered an extra \$40 (for a total of \$80).

We believe we have persuasive (but perhaps not conclusive) evidence that the higher amounts are more effective. Given the costs of continuing to run the experiment, we are seeking permission to end it, and simply offer \$80 to the small percentage of respondents who are interviewed in the last 2 weeks of each quarter (aka "the double-sample phase"). Ending the experiment was approved by the National Center for Health Statistics IRB on August 29, 2007.

THE DESIGN OF THE EXPERIMENT

In its "terms of clearance" on April 20, 2006, OMB said:

Approval is granted for the incentive plan as updated in a 4/12/06 email from Dr. Mosher. NCHS shall report to OMB via e-mail or in person the results of the incentive experiments as well as discuss changes needed over the clearance period.

Fieldwork for continuous interviewing in the NSFG is organized into 12-week quarters (four weeks each year are used for training and for holiday breaks). In each 12-week quarter, we offer a \$40 token in "Phase 1" (the first ten weeks of fieldwork). In "Phase 2," (weeks 11 and 12 of each quarter), we took a half-sample of the remaining cases (the "double sample") and the selected cases were divided into two groups:

- a) One group received **\$10** prepaid in addition to the standard \$40 (total \$50);
 - b) The other group received a **\$40** prepaid in addition to the standard \$40 (total \$80).
- (In either group, if the household had not completed a screener at the end of Phase 1, they were offered \$5 prepaid for the screener in Phase 2.)

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

In shorthand, the two approaches for the final two weeks of each quarter (“Phase 2”) were: \$5/~~\$10~~/\$40 versus \$5/~~\$40~~/\$40, where the key distinction is whether adults selected for the main interview were pre-paid \$10 or \$40.

The \$40 prepaid token was the approach taken in the last month of Cycle 6 fieldwork for about 724, or about 6% of our 12,571 respondents. Using this \$40 in addition to the standard \$40, we raised the weighted response rates from 64% to 79% while controlling fieldwork costs. We also showed a reduction in the survey’s bias, as these “late” respondents were significantly different from Phase 1 respondents with respect to marital status, Hispanic origin and race, and other key variables. (This is documented in the Series 1 and Series 2 reports from Cycle 6, which are posted on the NCHS web page at:

<http://www.cdc.gov/nchs/about/major/nsfg/nsfgcycle6reports.htm>

In our initial Cycle 7 protocol, we proposed to continue this Phase 2 “\$5/~~\$40~~/\$40” approach for the last 2 weeks of each quarter in Cycle 7 continuous interviewing. The NCHS Human Subjects Contact suggested that we had not demonstrated that a doubling of the main interview’s incentive (from \$40 to \$80) was necessary, even for this small group of respondents (typically less than 10%).

CYCLE 7 RESULTS, PART 1: RESPONSE RATES

Several complications in the administration of the double sample in the first quarter required us to exclude the results of Quarter 1 of year 1 in Cycle 7. (These were that some Phase 1 cases with appointments were carried over into Phase 2, reducing effort on the Phase 2 cases; the experimental procedures were confusing to some interviewers, and required additional training in Quarter 2; and the overall response rate was lower in that first quarter, as the entire data collection operation was being melded together.)

The table below presents the pooled results across Quarters 2 through 4. We show response rates separately (top & bottom panels) for household screener cases and main interview cases to be able to evaluate the potential impact of the incentives in each treatment group. Some sample cases were selected into Phase 2 at the screener stage, and some at the main interview stage.

Overall, 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 interview incentive. The response rates for main interview cases in Phase 2 were 12 percentage points higher in the higher-paid group.

Pooled Year 1, Quarters 2, 3 and 4 Unweighted Case Counts for Phase 2 Incentive Experiment Outcomes, Response Rates and Simple Random Sample Standard Errors

Screener Interview Cases in Phase 2							
	Sample Size	Completed Screeners	Refusals	Non-Interviews	Non-Sample	Response Rate	Standard Error
\$5/\$10/\$40	208	130	43	20	15	67%	3.4%
\$5/\$40/\$40	207	152	32	14	9	77%	3.0%

Main Interview Cases in Phase 2							
	Sample Size	Completed Main Interviews	Refusals	Non-Interviews		Response Rate	Standard Error
\$10 prepaid	192	100	48	44		52%	3.6%
\$40 prepaid	215	137	29	49		64%	3.3%

Note: 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.

RESULTS, PART 2: SAMPLE COMPOSITION

The table below shows the characteristics of the NSFG sample, in three categories:

Column (1) Phase 1 (weeks 1-10), offered \$40;

Column (2) Phase 2 (weeks 11-12), offered \$40 + \$10 = \$50; and

Column (4) Phase 2 (weeks 11-12), offered \$40 + \$40 = \$80.

	(1) Phase 1 \$40 N= 1,896	(2) Phase 2 \$50 N = 51	(3)= (1) vs. (2)	(4) Phase 2 \$80 N =68	(5)= (1) vs. (4)
FEMALE					
College Degree or more	34% (.01)	48% (.07)	NS	51% (.06)	.05
Ever had an abortion	6% (.01)	3% (.02)	NS	1% (.01)	.05
Ever had a live birth	59% (.01)	68% (.07)	NS	40% (.06)	.05
Ever had sex with female	13% (.01)	16% (.05)	NS	4% (.02)	.05
Income \$75,000+	17% (.01)	40% (.07)	.05	25% (.06)	NS
Multi-unit structure	38% (.01)	24% (.06)	.10	24% (.05)	.05
MALE	N=1,432	N=47		N=70	
Hispanic	20% (.01)	24% (.06)	NS	37% (.06)	.05
College Degree or more	28% (.01)	43% (.07)	.10	36% (.06)	NS
Ever fathered a birth	43% (.01)	48% (.07)	NS	36% (.06)	NS
Ever had sex with male	7% (.01)	5% (.03)	NS	1% (.01)	.05
Income \$75,000 +	25% (.01)	30% (.07)	NS	42% (.06)	.05
Multi-Unit Structure	37% (.01)	42% (.07)	NS	26% (.05)	.10
Physical Impediments	12% (.01)	10% (.04)	NS	20% (.05)	NS

The sample sizes in the experimental categories are small because only a sub-sample of the incomplete cases receives the “phase 2” protocol, and then that sub-sample is split into 2 incentive levels. The symbol “NS” after many of the entries in column (3) indicates that the difference between the \$40 group and the \$50 group is not significant.

But many of the differences between the \$40 group and the \$80 group are significant, using 2-tailed t-tests--suggesting that the \$80 incentive is getting different people into the sample, but the \$50 incentive is much less effective in that respect.

Given due caution about the sample sizes, however, some of the patterns are fairly clear and some are quite similar to those found in the final phase of Cycle 6:

- Women with college degrees were more likely to be in the \$80 group (51%) than in the Phase 1 \$40 group (34% had a college degree).
- Women who had had a live birth were less likely to be in the \$80 group (40%) than in the Phase 1 \$40 group (59%--in other words, childless women were more likely to be in the \$80 group. Since the principal outcome variable of the NSFG is fertility/birth rates, this is a critical finding).
- Women who had had sex with another woman were less likely to be in the \$80 group (only 4%) than in the \$40 group in Phase 1 (13%).
- For men, Hispanic men were more likely to be in the \$80 group (37%) than in the \$40 Group (20%). Given strong policy and program interest in this group, this is also a key finding.
- Men with college degrees appear to be more likely to be in the \$80 group (36%) than in the \$40 group (28%). This education difference is not significant with these sample sizes, but it is consistent with results for women in Cycle 7 and consistent with results for men in Cycle 6.
- For both men and women, those in the \$80 group were less likely to live in multi-unit structures than in the Phase 1 \$40 group (24% vs. 38% for women; 26% vs. 37% for men). In other words, those in the \$80 group were more likely to live in their own single-family homes.
- Finally, also for both men and women, those in the \$80 group were less likely to have had same-sex sexual contact than the \$40 group: 4% vs. 13% for women and 1% vs. 7% vs. 1% for men. Given the strong interests in these behaviors for public health reasons, these are also critical findings.

CONCLUSIONS

We have achieved consistent results in our response rates three quarters in a row, and that gives us some confidence that the results are robust. We therefore view the \$40 prepaid incentive (total of \$80) as preferable for our Phase 2 effort each quarter. We acknowledge

that based solely on simple-random-sample, classical hypothesis testing criteria, the results of the experiment are not based on sufficiently large sample sizes to clearly support choosing the \$40 prepaid incentive (using a 2-tailed test at the 5 percent level).

Our results comparing the sample composition of the two enhanced incentive levels with the standard \$40 amount are also less definitive than we would like because of small sample sizes, but many of the differences are significant, and they are broadly consistent with findings from Cycle 6 (2002 and 2003). The results suggest that busy, college-educated, childless, high-income people living in single-family homes are not as well represented in the standard Phase I sample as they should be. 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. We think this justifies the use of the \$80 amount.

We believe that there are good reasons to end the experiment now:

1. There are significant operational costs to the experiment. We experienced some logistical difficulties in the first quarter that damaged the results, we believe. Even in the later quarters, the experiment increases fieldwork costs due to requiring two different mailed correspondences, randomized assignment of segments, checking that interviewers applied the correct protocol, and analysis of the experimental data.
2. To the extent that one of the experimental groups achieves lower response rates or less representative sample composition, we are damaging the overall response rate of NSFG the longer the experiment continues.
3. While it is possible that the response rate advantage of the \$40/\$40 (\$80) incentive treatment will dissipate over time, it is very unlikely.
4. We estimate that it will 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. Extending the experiment, as we have pointed out, is expensive.

In sum, we recommend that we cease this experiment and use the \$40/\$40 (\$80) incentive to maximize response rates in the last two weeks of each quarter (Phase 2) for Year 2 of NSFG Cycle 7 and the foreseeable future. Approval to end this experiment was granted by the National Center for Health Statistics (NCHS) IRB—known at NCHS as the “Research Ethics Review Board”—on August 29, 2007 (Amendment 11, NCHS Protocol Number 2006-01).

REFERENCES**Cycle 5 Pretest Results**

A Duffer, J Lessler, et al. 1994. Effects of Incentive Payments on Response Rates and Field Costs in a Pretest of a National CAPI Survey. American Statistical Association (editor), 1994 Proceedings of the Section on Survey Research Methods, Volume II, pages 1386-1391.

W Mosher, W Pratt, and A Duffer. 1994. CAPI, Event Histories, and Incentives in the NSFG Cycle 5 Pretest. American Statistical Association (editor), 1994 Proceedings of the Section on Survey Research Methods, Volume I, pages 59-63.

JE Kelly, WD Mosher, et al. 1997. Plan and Operation of the 1995 National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 36, October 1997, pages 7-9, tables A and B. National Center for Health Statistics. Available on the NCHS web site at: <http://www.cdc.gov/nchs/products/pubs/pubd/series/ser.htm#sr1>

Cycle 6 Pretest and Main Study Results

RM Groves, G Benson, WD Mosher, et al. 2005. Plan and Operation of Cycle 6 of the National Survey of Family Growth. Vital and Health Statistics, Series 1, No. 42, August 2005. See pages 11-14, tables J, K, L, M on the Pretest; and pages 32-41 on the Main Study. National Center for Health Statistics. Available on the NCHS web site at: <http://www.cdc.gov/nchs/products/pubs/pubd/series/ser.htm#sr1>

W Mosher and Robert Groves. 2006. A Report to OMB on the National Survey of Family Growth: January 9, 2006. PowerPoint presentation given to OMB on January 9, 2006. See especially slides 18-20.

William D. Mosher, Ph.D., Project Officer, NSFG
NCHS
301-458-4385
e-mail: wdm1@cdc.gov

Robert M. Groves, Ph.D., Project Director,
ISR, University of Michigan
734-764-8365
e-mail: rgroves@umich.edu