MEMORANDUM

Date: January 22, 2015

To: Margo Schwab and Brian Harris-Kojetin
Office of Management and Budget

From: Joyce Abma (Contracting Officer's Representative) and Anjani Chandra (Team Leader, NSFG), Reproductive Statistics Branch, NCHS

Through: Amy Branum, Chief, Reproductive Statistics Branch, NCHS
Delton Atkinson, Director, Division of Vital Statistics, NCHS
Verita Buie, OMB Clearance Liaison, NCHS

Subject: **OMB Number 0920-0314**
Results of Incentive Experiment in 2013-2014 Continuous NSFG

BACKGROUND: EXPECTED FINDINGS

The National Survey of Family Growth (NSFG) has been experimenting with different incentive amounts in Phase 1 since September 2013. The experiment randomized area segments to two different treatments (the current $40 or an experimental amount of $60). The Phase 2 incentive was the same for both arms: all respondents received $80 for completing the interview. We have now completed four quarters of data collection under this experimental design.

At the beginning of the experiment, we hypothesized that this incentive change would increase response rates approximately 7%. We also expected there to be savings in effort from an increased incentive such that the incentive would be, at least, cost neutral. At the time we proposed the experiment, we observed that interviewers were making more calls to produce fewer interviews. We presented the following information:

	2006-2010	**2011-2012**
Hours of labor per interview:	9.1 hours	9.5 hours
Average no. of visits to HH per interview:	7.1 visits	8.1 visits
Response rate:	76.6%	72.7%

It was our hope that an increased incentive would increase both efficiency and response rates.

Finally, with respect to who would be recruited by the new protocol, we noted that during the current cycle, we had continued to be very successful at recruiting Hispanic respondents. However, response rates for White/Other and Black respondents, particularly males, had fallen compared to our experience in a previous cycle.

	2006-2010	**2011-2012**
Male	**75%**	**72%**
Black	**77%**	**73%**
Hispanic	74%	73%
White & other	**75%**	**72%**
Female	**78%**	**73%**
Black	81%	77%
Hispanic	80%	79%
White & Other	**76%**	**70%**

To return to our hypotheses, we expected that a $20 increase in the Phase 1 incentive would produce a 7-8 percentage point increase in the response rate to the survey, and result in lower interviewer labor costs per case, sufficient to pay for most of the cost of the incentive increase. Given that the Phase 2 incentive was not changed, we expected most of these gains to come in Phase 1 of NSFG data collection.

BASIC NSFG PROCEDURES BEFORE THE INCENTIVE EXPERIMENT

Briefly, the NSFG is conducted in person using an area probability sample. About 20,000 addresses are selected each year in about 35 Primary Sampling Units (PSUs) across the nation. The interviewing process is divided into "stages." The first stage involves identifying eligible persons using a screening interview that produces a list of all the persons in a household, including their age, sex, race, and ethnicity. About 14,000 "screeners" are completed each year to identify households containing the NSFG-eligible population, persons 15-44 years of age. About half of those households contain at least one person 15-44. In the second stage, a randomly selected person is given a "main" interview that obtains the substantive content of the NSFG.

The fieldwork is organized into four 12-week "Quarters" (abbreviated Q below). Each 12-week Quarter is divided into 2 "Phases," following responsive design principles:

- Weeks 1-10 comprise Phase 1, in which a $40 conditional incentive is offered, and interviewers are given a larger caseload. By the end of this Phase, the response rate is about 55% and we have expended about 8 hours of interviewer labor per case.

- Weeks 11-12 comprise Phase 2, in which 2 major changes are made: first, a one-third sample is taken of the remaining non-responding cases, thereby lowering the caseload from 45% to 15% of the original sample – and allowing interviewers to spend 3 times as many hours per case as they did in Phase 1; and second, the conditional incentive is increased to $80.

The main interview response rate in Phase 1 was about 58% in 2006-2010 (Lepkowski et al, Series 2, No. 158, table N), but it has been falling to closer to 50% in recent quarters. This trend

raised concerns about nonresponse bias as well as inadequate sample yield – both for the entire population 15-44 and in some key population sub-groups.

EXPERIMENTAL PROCEDURES

Key elements of the experimental design are as follows:

- The incentive was either $40 (control) or $60 (experimental) in Phase 1 (weeks 1-10) of each quarter.
- Both Phase 1 groups were offered the same incentive ($80) in Phase 2.

Randomization to experimental and control conditions for Phase 1 occurred at the segment level. This has more statistical power than randomizing at the PSU level, but it does require each interviewer to use different procedures in the segments she is assigned each quarter (each interviewer is assigned 2-3 segments each quarter; all interviewers take part in the experiment). Randomizing at the sample line (household or respondent) level would have offered still more statistical power, but it would also increase the odds of interviewers inadvertently offering the wrong incentive to a sampled person, and increase the risk that neighbors would discover that others are being offered different amounts.

RESULTS OF THE EXPERIMENT

Since the experimental treatment was applied to Phase 1, we look at the results first at the end of Phase 1. We will also examine the results from Phase 2 and cumulatively since the experimental treatment may have "carryover" effects.

1. **Phase 1 Case Counts and Response Rates**

The first panel of Table 1 gives sample sizes, completed screening interviews, and completed main interviews by the two treatments as they stood at the end of Phase 1. The second panel presents unweighted screener and main completion rates, as well as overall response rates for Phase 1. The counts in Tables 1 and 2 are different. The counts in Table 1 are sample sizes, completed screeners and completed main interviews. These are the numerators in the various response rate calculations. The N's in Table 2, on the other hand, are denominators of the response rate – the total sample size minus nonsample cases.

Table 1. Phase 1: Unweighted Case Counts and Completion and Response rates

Unweighted Phase 1 Case Counts		
	$40	**$60**
Sample Size	9,926	9,933
Completed Screeners	7,273	7,304
Completed Main Interviews	2,173	2,495

Screener and Main Completion Rates and Combined Response Rates						
	$40			**$60**		
	N	%	SE(%)	N	%	SE(%)
Screener Completion Rates	8,587	84.7%	0.4%	8,747	83.5%	0.4%
Main Completion Rates	3,494	62.2%	0.8%	3,728	66.9%	0.8%
Overall Response Rates		52.7%			55.9%	
Eligibility Rate	7,272	48.0%	0.6%	7,304	51.4%	0.6%

As shown in the first panel, the $60 treatment appeared to increase the production of completed screening interviews slightly. It led to a more substantial increase in the number of completed main interviews. We will examine this in more detail later.

As shown in the second panel, the screener response rates are approximately the same across the two treatment groups. It does not appear that the incentive is having an effect on this stage of the process. The main completion rates, however, are significantly different. Here, the $60 treatment increases the rate at which main interviews are completed (p=0.0041), a difference of 4.7 percentage points. The overall (unweighted) response rates at the end of Phase 1 are listed in the bottom row. These are not as far apart as hypothesized. The overall Phase 1 response rate for the $60 treatment is about 3.2 percentage points higher than that for the $40 control group. Finally, the eligibility rates are significantly higher for the higher incentive group.

Table 2 shows interview rates and completion rates for several key demographic subgroups at the end of Phase 1.

Table 2. Phase 1 Main Interview Completion Rates by Demographic Subgroups

	$40			$60			Comp Rate Diff ($60-$40)	*p-value*
	Eligible Persons (N)	**% Complete**	***(SE)***	**Eligible Persons (N)**	**% Complete**	***(SE)***		
Race								
Hispanic	770	61.9%	*1.8%*	867	69.1%	*1.6%*	7.2%	0.017
Black	659	68.7%	*1.8%*	667	74.4%	*1.7%*	5.7%	0.138
White & Other	2065	60.2%	*1.1%*	2194	63.8%	*1.0%*	3.6%	0.004
Gender								
Male	1642	59.2%	*1.2%*	1713	64.7%	*1.2%*	5.5%	0.008
Female	1852	64.8%	*1.1%*	2015	68.8%	*1.0%*	4.0%	0.036
Age								
15 - 19	635	72.3%	*1.8%*	694	77.1%	*1.6%*	4.8%	0.058
20 - 44	2859	60.0%	*0.9%*	3034	64.6%	*0.9%*	4.6%	0.012

Table 2 shows that the $60 treatment produced higher completion rates for every subgroup. The range of subgroup response rates is higher for the $40 treatment (72.3-59.2=13.1%) than the $60 treatment (77.1-63.8=7.3%). Reducing the variance of these subgroup response rates may be useful in controlling nonresponse bias, but at the very least is useful for controlling potential nonresponse adjustments. The biggest increases did not occur for the "White and Other" and "Black" subgroups – the groups that had seen larger declines in response rates before the experiment began – as we had expected. Increasing main interview rates for these groups will be a continued focus of the NSFG team.

Thus, as shown by Tables 1 and 2: at the end of Phase 1, we observed the hypothesized increase in (unweighted) response rates (Table 1, 2nd panel). We found that these rates increased for all subgroups and that the variation in the rates was somewhat decreased (Table 2). However, the increase in the Phase 1 response rate was not as large as expected.

2. Phase 2 and Cumulative Case Counts and Response Rates

We now turn our attention to the Phase 2 and cumulative results. During Phase 2, all respondents received the same incentive for completing the interview ($80).
Table 3 shows the Phase 2 and the overall (Phase 1 and 2 combined) counts of completed screener and main interviews. The differences between the two treatments, in terms of counts of screener and main interviews, are small. See Appendix I for hypotheses and discussion of this lack of difference.

Table 3: Unweighted Case Counts		
	$40	$60
Completed Screeners		
Phase 2	218	237
Overall	7,491	7,541
Completed Main Interviews		
Phase 2	262	230
Overall	2,435	2,725

Table 4 provides the results that are parallel to those presented in Table 2. The weighted Phase 2 screening response rate was 57.1% for the $40 group and 56.0% for the $60 group. This difference is not significant. Table 6 also shows the unweighted and weighted final screening and main interview rates, and a final combined response rate. Finally, we note that the incentive seemed to produce a higher eligibility rate.

Table 4: Screener and Main Completion Rates and Combined Response Rates						
	$40			$60		
	N	%	SE(%)	N	%	SE(%)
Screener Completion Rates						
Weighted Phase 2 Rate	344	57.1%	2.7%	369	56.0%	2.6%
Unweighted Final Rate	7,010	96.2%	0.2%	7,148	95.9%	0.2%
Weighted Final Rate	7,010	92.8%	0.3%	7,148	92.3%	30.0%
Main Completion Rates						
Weighted Phase 2 Rate	457	54.8%	2.3%	481	46.7%	2.3%
Unweighted Final Rate	2,850	85.4%	0.7%	3,177	85.8%	0.6%
Weighted Final Rate	2,850	77.7%	0.8%	3,177	77.0%	0.7%
Overall Response Rates						
Weighted Phase 2 Rate		31.3%			26.2%	
Unweighted Final Rate		82.2%			82.3%	
Weighted Final Rate		72.1%			71.1%	
Weighted Final Eligibility Rate	6,741	47.7%	0.6%	6,856	51.4%	0.6%

It appears that Phase 2 was somewhat less effective for the $60 treatment. As a result, the final response rates are very similar across the two treatments. Thus, it appears that, with respect to response rates, the impact of the second Phase is somewhat mitigated following the $60 treatment in the first Phase.

Table 5 shows the weighted Phase 2 main interview rates by demographic subgroup. In this case, there is a consistent pattern that favors the $40 treatment. In each case, the $40 treatment has the higher main interview rate. Several of these differences are significant or marginally significant.

Table 5. Weighted Phase 2 Main Interview Rate by Demographic Subgroup

	$40			$60			Comp Rate Diff ($60-$40)	*p-value*
	N	% Complete	*(SE)*	N	% Complete	*(SE)*		
Overall								
	457	54.80%	*2.30%*	481	46.70%	*2.30%*	*-8.10%*	0.069
Race								
Hispanic	94	60.00%	*5.10%*	100	56.50%	*5.00%*	*-3.50%*	0.545
Black	81	60.30%	*5.50%*	69	52.10%	*6.10%*	*-8.20%*	0.403
White & Other	282	51.20%	*3.00%*	312	42.10%	*2.80%*	*-9.10%*	0.049
Gender								
Male	226	50.90%	*3.30%*	226	48.60%	*3.30%*	*-2.30%*	0.682
Female	231	58.40%	*3.20%*	255	45.00%	*3.10%*	*-13.40%*	0.026
Age								
15 - 19	59	68.00%	*6.10%*	47	43.90%	*7.30%*	*-24.10%*	0.028
20 - 44	398	52.90%	*2.50%*	434	47.00%	*2.40%*	*-5.90%*	0.186

Table 6 shows the cumulative (Phase 1 and 2 combined) weighted main interview completion rates for the same demographic subgroups. There is no clear pattern favoring either treatment and none of the differences are significant. It appears that final response rates are not differentiated between the two treatments for these demographic subgroups.

Table 6. Cumulative Weighted Main Interview Completion Rates by Demographic Subgroups

	$40			$60			Comp Rate Diff ($60-$40)	*p-value*
	Eligible Persons (N)	% Complete	*(SE)*	Eligible Persons (N)	% Complete	*(SE)*		
Race								
Hispanic	621	80.30%	*1.60%*	737	82.00%	*1.40%*	1.70%	0.463
Black	562	82.10%	*1.60%*	580	84.30%	*1.50%*	2.20%	0.412
White & Other	1667	75.20%	*1.10%*	1860	72.90%	*1.00%*	-2.30%	0.214
Gender								
Male	1315	74.90%	*1.20%*	1433	76.80%	*1.10%*	1.90%	0.501
Female	1535	80.10%	*1.00%*	1744	77.20%	*1.00%*	-2.90%	0.117
Age								
15 - 19	599	81.40%	*1.60%*	648	81.40%	*1.50%*	0.00%	0.984
20 - 44	2251	76.80%	*0.90%*	2529	76.10%	*0.80%*	-0.70%	0.648

In sum: Regarding response rates and interview counts, the $60 treatment appears to improve both during Phase 1. Unfortunately, the gains made in Phase 1 are reversed in Phase 2. We speculate that this is due to the decreased effectiveness of the increased incentive (an increase

of $20 vs $40). However, the gains in the number of interviews from Phase 1 are retained. The net result is that a larger number of interviews are completed under the $60 treatment.

3. Effort indicators

In this section, we explore factors associated with costs. In a face-to-face survey, cost comparisons between cases assigned to the same interviewer are difficult. Interviewers generally report their total time without assigning that time to specific cases. Therefore, we are required to use indirect measures or model-based estimates.

Table 7 presents several indirect measures of effort, for Phase 1 and overall (Phase 1 and Phase 2 combined). The first is the number of call (in-person visits) attempts per completed interview. This includes all call attempts, even those made to cases that prove to be ineligible or refuse to complete the main interview. The number of call attempts is smaller for the $60 group, particularly during Phase 1, where the $40 group required 3.31 (15%) more calls per interview. Of course, not all calls are the same length. The composition of the calls also matters. We will examine this later using a model-based approach.

Table 7: Selected Effort Indicators			
	$40	$60	Difference
Calls (in-person visits) per completed interview			
Phase 1	24.71	21.40	3.3
Overall	23.98	22.59	1.4
Contacts per Main Interview			
Phase 1	7.75	6.96	.8
Overall	7.25	6.89	.4
Resistant Attempts per Completed Interview			
Phase 1	0.91	0.82	.09
Overall	0.95	0.95	.00

An additional indicator is the number of contacts (speaking with a household member) per main interview. Contacts generally require more time than non-contacts. One outcome for the incentive would be to reduce the number of contacts required to complete the interview. For this measure, the $60 treatment reduces the number of required contacts to complete an interview. Again, the difference is smaller after Phase 2 (“Overall”, in the table).

The final measure is the number of resistant attempts per completed interview. These resistant attempts require further effort to address respondent concerns. For this indicator, the $60 treatment has some impact in Phase 1, but the gains are lost by the end of Phase 2, as seen in the “overall” row.

One way to estimate the cost savings is to assume that each call takes about the same amount of time. That is, we can divide all the interviewer hours by the total number of calls made. Using this approach, a call takes about 25 minutes on average. Therefore, a reduction of 1.39 calls per

interview yields an estimated savings of about 34.8 minutes per interview. This is likely an overestimate of the savings as the calls saved are generally shorter (i.e., not interviews).

We also used a regression model approach, where the number of calls of different types is used to predict the hours worked each week for each interviewer, to account for the differences in length of call. Estimates from this regression model lead us to predict savings of about 0.8 hours per interview under the $60 treatment.

An interviewer hour (excluding recruitment, hiring, and training costs) costs about $35. Therefore, we expect the additional incentive cost of $20 to produce savings between $20.30 and $28.00 of interviewer time per interview. This is a net savings of between $0.30 and $8.00 after factoring in the additional cost of the higher incentive. These estimates are based on experimental data. The savings might be higher or lower when interviewers no longer have sample available to work with both treatments ($40 and $60). Overall, the cost savings appear to be minimal, but the incentive does at least pay for itself.

4. Assessment of Potential Nonresponse Biases

In this final section, we assess the impact of the incentive on any potential nonresponse biases. These assessments are an additional leg of the analysis in a total survey error perspective.-

One approach for assessing nonresponse bias involves comparing percentages of the two experimental samples that are in subgroups of key relevance to the substance and mission of the survey, and/or are hard to reach or less likely to agree to an interview, as suggested by knowledge gained from fieldwork experience. For example, the percent with a four-year college degree might be increased with a higher incentive as this group is likely to have higher incomes and may be less likely to respond for lower incentive amounts. Admittedly this differs from an approach that would compare these percentages to their levels in the population, to see which group comes closer. However if enough evidence is found that the incentive amount results in significantly different percentages in the key categories, such comparison (with actual population levels) could then be pursued, where possible.

Table 8 presents these percentages, or “key estimates”, for female and male respondents aged 15-44.

The last column in the table is a p-value for the test that has the null hypothesis of no difference in the means.

Table 8: Selected Key Statistics by Sex					
Variable Name	$40		$60		
	N	Statistic	N	Statistic	*p-value*
Females					
Percent with 4-yr college degree	1,338	33%	1,501	30%	0.30
Percent with income $100,000+	1,224	13%	1,372	14%	0.70
Percent never married	1,340	47%	1,503	45%	0.49
Percent ever cohabited	1,340	55%	1,503	57%	0.52
Percent ever had sex in past 12 months	995	88%	1,105	89%	0.90
Percent ever had an abortion	817	19%	897	20%	0.47
Percent ever had female-female sex	1,332	18%	1,493	18%	1.00
Percent with no births	1,340	45%	1,502	47%	0.30
Males					
Percent with 4-yr college degree	1,091	21%	1219	29%	0.03
Percent with income $100,000+	963	17%	1094	17%	0.98
Percent never married	1,092	57%	1221	52%	0.12
Percent ever cohabited	1,092	35%	1221	32%	0.24
Percent ever had sex in past 12 months	908	89%	987	90%	0.48
Percent used a contraceptive method at most recent sex	517	88%	518	84%	0.05
Percent who did not father a birth	1,092	63%	1,221	62%	0.62
Percent with co-residential kids	908	36%	987	37%	0.29

In Table 8, there are no significant differences for the female variables. Two of the male variables show significant differences – “Percent with 4-yr college degree” and “Percent used a contraceptive method at most recent sex.”

The general pattern is one of no differences. This result can be interpreted as meaning that the increased Phase 1 incentive did not lead to either an increase or decrease in nonresponse bias.

DISCUSSION

The results of the incentive experiment are mixed. The expected increase in response rate was not realized. However, the incentive increase did have an impact on Phase 1 by increasing the response rates in that Phase. This led to a higher overall yield.

Unfortunately, this also seemed to reduce the effectiveness of an $80 incentive in Phase 2. This meant that many of the gains made from using the $60 incentive were lost during Phase 2. The result was that the final response rates, subgroup response rates, and key estimates were quite similar across the two treatment groups.

Our cost comparisons show that even after Phase 2, there were still small cost savings that exceeded the expense of the incentive, or at least the incentive covered its own cost. Therefore, the larger incentive appears to reduce "total survey error" by reducing sampling error (expressed as either higher yield or cost savings) without increasing nonresponse error.

In sum: Overall, the $60 incentive does not appear to increase overall response rates. Nor does it lead to changes in estimates that might be interpreted as reductions in nonresponse bias relative to the $40 incentive. There is, however, a somewhat larger yield for the $60 incentive, which may contribute to reductions in sampling error and data collection costs.

RECOMMENDATION

We find insufficient evidence to propose an increase in the Phase 1 incentive to $60, and propose to revert to the incentive structure in place prior to the survey: $40 incentive in Phase 1 and $80 in Phase 2. We will continue to explore other ways of increasing yield, reducing nonresponse bias, and containing costs in the NSFG.

REFERENCES

Axinn WG, CF Link, RM Groves. 2011. Responsive survey design, demographic data collection, and models of demographic behavior. Demography 48(3):1127-1149.

Duffer A, J Lessler, M Weeks, W Mosher. 1994. Effects of Incentive Payments on Response Rates and Field Costs in a Pretest of a National CAPI Survey. 1994 Proceedings of the Section on Survey Research Methods, Volume II: Papers presented at the 49th Annual Conference of the AAPOR, 1994. Pages 1386-1391.

Groves R, R Cialdini, MP Couper. 1992. Understanding the Decision to Participate in a Survey. Public Opinion Quarterly 56: 475-495.

Groves R, E Singer, A Corning. 2000. Leverage-Saliency Theory of Survey Participation. Public Opinion Quarterly 64(3): 299-308.

Groves R, S Presser, S Dipko. 2004. The Role of Topic Interest in Survey Participation Decisions. Public Opinion Quarterly 68 (1): 2-31.

Groves R, G Benson, W Mosher, et al. 2005. Plan and Operation of the 2002 National Survey of Family Growth. **Vital and Health Statistics, Series 1, No. 42**. August 2005.

Kelley JE, W Mosher, A Duffer, S Kinsey. 1997. Plan and Operation of the 1995 National Survey of Family Growth. **Vital and Health Statistics, Series 1, No. 36.** October 1997. See pages 1-9.

Lepkowski J, W Mosher, R Groves, et al. 2013. Responsive design, Weighting, and Variance Estimation in the 2006-2010 National Survey of Family Growth. **Vital and Health Statistics, Series 2, No. 158.** June 2013. See especially Appendix II.

Mosher W, W Pratt, A Duffer. 1994. CAPI, Event Histories, and Incentives in the NSFG Cycle 5 Pretest. 1994 Proceedings of the Section on Survey Research Methods, Volume I: Papers presented at the Annual Meeting of the American Statistical Association, Toronto, Canada, August 1994. Pages 59-63.

Peytchev A, Peytcheva A, R Groves. 2010. Measurement Error, Unit Nonresponse, and Self-reports of Abortion Experiences. Public Opinion Quarterly 74 (2): 319-327.

Appendix I
Understanding Phase 2 and Cumulative (Phase 1 + Phase 2) results: Hypotheses

The Phase 2 samples for the two treatments could differ in at least two important ways:

1. The $40 treatment had lower response rates at the end of Phase 1. We might expect the Phase 2 sample in the $60 Phase 1 arm to be more difficult to interview. This might be reflected by the Phase 2 sample for that arm having somewhat lower estimated response propensities than the $40 arm. If the Phase 2 samples differ in this way, then we might have different results at the end of Phase 2 with respect to response rates.

2. The increase in the amount of the incentive was different for the two groups. The $60 treatment group had its incentive increased by $20 for Phase 2, while the $40 treatment group had their incentive increased by $40. We might expect the smaller increase to have a relatively smaller effect on final response rates.

We can evaluate the first point by examining the estimated response propensities for the Phase 2 samples by treatment. The treatment indicator was not used as a predictor in the propensity models. The models include a large set of fixed, baseline characteristics as well as elements from the time-varying paradata (e.g., contact observations, number of calls, number of contacts, ever resistant, etc.). Table 4 shows the mean of the estimated response propensities for the Phase 2 samples by current status (unscreened and screened) and treatment ($40 vs $60). The estimated propensities are actually higher for the $60 treatment, although not significantly so. Thus, it does not appear that sample quality for Phase 2 favors the $40 treatment.

Table 4: Phase 2 Sample Mean of the Estimated Response Propensities and Weight Variance

	$40	$60
Screener	0.151 *(0.005)*	0.157 *(0.005)*
Main	0.062 *(0.005)*	0.068 *(0.005)*
Mean Phase 2 Weight (SE)	3.01 (0.03)	2.99 (0.03)
1+L	1.098	1.095

We also examined the second Phase weights to see if the experiment created any imbalances in the weights. If one treatment finished Phase 1 at a higher response rate, it may have required a larger portion of the remaining cases to be sampled for Phase 2. This would lead to a reduction in the weighting factors for this Phase of sampling. This could be a positive outcome since lower variability in the weights may lead to lower variance estimates. However, the results in Table 4 show that the weights, and their expected maximal influence on variance estimates ("1+L"), are roughly identical across the two treatments.

Of course, the experimental condition may have some influence on the sampling. The second Phase sample is drawn to create a sample of size sufficient for an interviewer with a specified

level of productivity who works 30 hours per week. In this design, the sampling rates for the second Phase in one segment are linked to the sampling rates in another. This situation may push the sample to an overall average rate. If only one treatment was being used, different rates might result.

Hypothesis 2 (that the smaller increase in Phase 2 incentive from $60 to $80 is less effective) cannot be evaluated with our current data. The experiment was not designed to evaluate interactions with the Phase 2 design. We could imagine further experimentation to see if the performance of Phase 2 could be improved (see the Discussion section).