Data That Do Not Persuade: Results of an Incentive Experiment

Ronald S. Fecso (NSF)
Ronald J. Broach (NORC)
Karen H. Grigorian (NORC)

National Science Foundation, Division of Science Resources Statistics
National Opinion Research Center

rfecso@nsf.gov
Broach-Ronald@norc.uchicago.edu
Grigorian-Karen@norc.uchicago.edu

Abstract

In the current environment of declining response rates, many survey researchers debate whether to include an incentive to increase response. The questions are: What type of incentive to use? How effective will that incentive be and how much increase in response rate will be gained? While many opinions about incentives exist, and those that do use an incentive generally measure response rate impact, it is rarer to find a scientific experiment that evaluates the effect of the incentive on both response rate and data quality. Controlled incentive experiments were conducted on the 2003 Scientists and Engineers Statistical Data System (SESTAT) surveys. The incentives varied in each of three related surveys. In the Survey of Doctorate Recipients (SDR) a $30 prepaid incentive, a $50 postpaid incentive, and an NSF Brochure with data CD incentive were used as refusal conversion tools late in the data collection period. In The National Survey of College Graduates (NSCG), a brochure experiment was embedded into the first mailing. The National Survey of Recent College Graduates (NSRCG) included a brochure in an experimental treatment.

In this paper, we focus only on a quantitative and qualitative evaluation of the effect of using the non-monetary incentive --brochures. This paper demonstrates that using a brochure had a negative impact on both item response rates and data quality. These findings are significant.

Keywords: Response Rate; Non-monetary Incentive; Brochures; SESTAT; Survey of Doctorate Recipients; Data Quality, Item Nonresponse; Data Editing, Cost

Introduction and Background

The purpose of the National Science Foundation's (NSF) Scientists and Engineers Statistical Data System (SESTAT) is to provide information on the entire U.S. population of scientists and engineers with at least a bachelor's degree. SESTAT is produced by combining data from the Survey of Doctorate Recipients (SDR) (representing persons in the general U.S. population who have earned a doctorate in science or engineering from a U.S. institution), the National Survey of Recent College Graduates (NSRCG) (representing persons with a recently earned bachelor's or master's degree in science and engineering from a U.S. institution) and the National Survey of College Graduates (NSCG) (representing, in the 2003 survey, all individuals in the U.S. at the time of the decennial census with a bachelor's degree or higher).

Historically, response rates for the SESTAT surveys are around 80%. The 2003 surveys, like many surveys, faced numerous challenges to achieving this response rate. In an effort to explore potential methods that might help slow the decline in response, NSF worked with the survey contractors to embed several incentive experiments into the surveys. In the NSCG and NSRCG, brochure experiments were embedded into the initial mailings. In the SDR, the experiment was developed in response to circumstances. As we neared the end of the field period for the SDR, there were quite a number of sample members who were not locating problems but who had not completed the survey in spite of attempts to persuade them via mail, telephone and e-mail. In conjunction with NSF, National Opinion Research Center (NORC) staff developed and fielded a controlled experiment to determine whether a Brochure incentive would boost our response rates. In the SDR, we also studied the impact on data quality.

Previous research has clearly demonstrated that incentives increase response rates on mail and telephone surveys (Church, 1993; Fecso, 2001; Fox, Crask, and Kim, 1988; Gunn and Rhodes, 1981; Harvey, 1987; Hopkins and Gullickson, 1992; Kovac and Markesich, 2002; O'Brien, Levin, Hagerty-Heller, Greenlees, and Kirsch, 2001; Singer, Groves, and Corning, 1999; Tzamourani and Lynn, 1999), although the cost effectiveness is not clearly demonstrated. In the brochure experiments, the cost of the brochure is very small compared to other survey costs. We anticipated that incentives offered to this group of nonrespondents would elicit an increased response rate over and above a simple final appeal. Indeed, this was the case.

An important aspect of cost-effectiveness is data quality. Research on the quality of data obtained from the use of incentives is less common and seems to focus on item nonresponse as the measure of data quality. Tzamourani and Lynn (1999) indicate that no difference was found in the item response rate between incentive and non-incentive groups. Datta, Horrigan, and Walker (2001) also indicate that there is no statistically significant impact on item nonresponse between incentive and non-incentive groups.

In this paper, we offer a quantitative and qualitative evaluation of the effect of using a high quality NSF Brochure including one test with a data CD as an incentive. As a quantitative measure, we provide response rate information for the NSCG and SDR surveys. As a very interesting twist to such experiments, the SDR study provides a qualitative comparison of cases from the SDR Brochure treatment group compared to the control. The qualitative comparison is defined as the amount of data editing and cleaning required on the completed surveys, and an evaluation of the amount of item non-response as well as the quality of data at selected key data items. We analyzed subgroups of cases based upon sex, race/ethnicity, citizenship status at birth, past response to the survey, and age. The SDR incentive experiment was implemented late in data collection, and so all members of the experiment and control groups were non-respondents at the time of the treatment implementation. All evaluated cases had already received at least one mail questionnaire, access to the Web-based version of the questionnaire, and telephone contact from interviewing staff. Additionally, we provide a cost comparison of the Brochure incentive in SDR. We conclude with our overall evaluation of the Brochure experiments for both the NSCG and SDR surveys.

Generally, we find that the Brochure incentive did not have a significant impact on response rate for this population. We employ a multifaceted approach to measuring data quality, one that goes beyond simple item nonresponse. We find, contrary to other researchers, that there are differences in data quality between the incentive and non-incentive groups with the Brochure incentive groups providing less data and lower quality data than the control group. The Brochure incentive also added some cost above that of the control treatment. Overall our findings do not support the use of a Brochure as an incentive for refusal conversion late in the data collection period.

SDR Survey Background

Background and Purpose of the SDR

The Survey of Doctorate Recipients is a longitudinal panel survey of individuals who have received doctoral-level degrees from U.S. institutions in science and engineering (S&E) fields and are pursuing their professional careers in the United States. The goal of the SDR is to provide policymakers and researchers with high-quality data and analyses for making informed decisions related to the education achievement and career movement of the Nation's doctoral scientists and engineers. The U.S. Congress, several federal agencies, researchers, professional associations, and academic institutions use the results of this survey.

The SDR has been conducted biennially since 1973 for the National Science Foundation (NSF) in conjunction with the National Institutes of Health and other sponsors. In this survey, a sample of holders of S&E doctorates earned at U.S. institutions is followed throughout their careers from year of doctorate degree award through age 75. Every two years, a sample of new S&E doctoral degree earners is added to the SDR sampling frame from another NSF-sponsored survey, the Survey of Earned Doctorates. For the 2003 cycle, a sample of individuals under the age of 76 who have earned doctoral degrees in science and engineering from U.S. institutions through academic year 2002 was surveyed.

For further details regarding the 2003 SDR refer to the *2003 Survey of Doctorate Recipients Methodology Report*.

Methods

Data Collection Protocol

In May 2004, after seven months of data collection, non-respondents had been sent up to four mail contacts (Advance Letter, first questionnaire mailing, Thank You/Reminder Postcard, second questionnaire mailing), had been sent between one and three e-mail contacts, and CATI interviewers had made up to 20 attempts to contact the sample member by telephone. At the time this Incentive Experiment started, 71% of the sample had responded to the survey. The majority of the non-respondents remaining at this stage in the data collection were considered the "most difficult" portion of the sample since they had not responded to multiple contact attempts through mail, telephone and/or E-mails during the previous 7 months.

In the experiment, we mailed the 2004 Science & Engineering Indicators (SEI) Information Card to sample members in the Brochure incentive treatment group. The Information Card is a high-quality, professionally-produced four color bifold brochure with pockets that included summary tables from the 2004 SEI report as well as a CD containing additional summary information. (PDFs of the brochures are included at the end of the paper.) Along with this Brochure, a gaining cooperation letter explained that the Brochure contained reports derived from SDR data and requested participation in the current SDR. The letter contained access information for the 2003 SDR Web survey as well as the toll-free number to complete the survey by telephone.

Simultaneous with the USPS mailing, an E-mail message was sent containing the same information and direct links to access the online web survey.

The primary metrics used to measure quality for this experiment were overall response and cooperation rates. In addition to overall response, NORC developed three other metrics to assess the quality of the data resulting from the completed cases in this experiment. These data quality metrics are described briefly below.

Data Quality Metric—Imputation/Item Nonresponse Flag Score

The first metric is a simple measure of item nonresponse. The SDR data were subject to editing for range compliance, for internal consistency, to correct skip pattern errors, and so forth. Any variables that were missing data after this editing would be imputed. So the first measure of data quality is the Imputation Flag Score.

There are 257 variables available for imputation in each case. For the purposes of creating the Imputation Flag Score, we have restricted the calculation to 224 variables. We did not include variables in the calculation that are sample management variables that were not collected during the 2003 SDR.

The denominator is the 224 variables available for imputation included in this analysis. To create the numerator, we counted the total number of variables where the imputation flag was set to '1'.

The Imputation Flag Score calculation can be expressed with the following formula…

Imputation Flag Score = (Count of Imputed Variables / Count of Evaluated Variables) x 100

The higher the value, the more variables were imputed for that case. A low value indicates a low item non-response rate.

Data Quality Metric—Edit Flag Score

Beyond this, we evaluated the percentage of variables in a case that were changed (and flagged as changed) by the edit and cleaning rules. This Edit Flag Score is an indicator of the level of effort needed to clean the data in a case so that it meets project standards. Since all cases were subject to the same Edit steps, this metric provides a reliable measure of the quality of the data for each case.

There are up to 339 variables available for editing in each case. For the purposes of creating the Edit Flag Score, we have restricted the calculation to 253 variables, the denominator for the calculation below. We did not include variables in the calculation that are coded verbatim texts, self-codes, or sample management variables that were not collected during the 2003 SDR.

The following calculation is called the Edit Flag Score:

Edit Flag Score = (Count of Edited Variables / Count of Evaluated Variables) x 100

For this calculation, we included the same 253 variables as above.

To create the numerator, we evaluated the six edit flags for each variable. If any of the edit flags were set to '1', we count this as an edited variable. The total count of the edited variables for a case is the numerator for the calculation.

The higher the value, the more variables were edited for that case. A low value indicates less editing or data cleaning was needed for the case and this indicates that the data quality is higher when the Edit Flag Score is low.

Data Quality Metric—Critical Item Score

An additional Critical Item Score was calculated on a subset of 25 variables that are identified by NSF as required for completeness (n=6) or highly desirable for analysis (n=19).

The Critical Item Score is based on just these critical items. This score is calculated like the Edit Flag Score with the exception that it is restricted to the critical items.

The higher the value, the more variables were edited for that case. As with the Edit Flag Score, a low value indicates less editing or data cleaning was needed for the case and this indicates that the data quality is higher when the Edit Flag Score is low.[*]

Tables and Analysis

First, we examined overall response and cooperation rates for the experiment and control groups. Table 1 provides a detailed look at the response rate for each of the groups in the Incentive experiment.

From Table 1, we see that the control group and the Brochure Incentive treatment group responded at statistically similar rates. Even with subgroup sample sizes that were quite small, the rates were usually similar for practical purposes as well.

Next we examined the data quality metrics. Table 2 shows the results of comparing the treatment and control groups on the three data quality metrics described above.

There is a clear increase in item nonresponse (or decrease in item response) in the treatment group over the control group (p=.0105). This indicates that the control group members who responded provided data for more questions than the responders from the Brochure incentive group. The mean number of variables with no data from the respondent for the Control group was 12.0; the mean number of variables with no data from the respondent for the Brochure incentive group was 17.9. The differences were significant for the young and male subdomains.

We see a negative impact on data quality in the Edit Flag scores. Both Edit Scores show clear increases in the data editing required to bring the data from the treatment group up to SESTAT standards when compared with the data from the control group. The mean number of variables requiring editing for the Control group was 6.5; the mean number of variables requiring editing from the Brochure incentive group was 10.2 (p=.0012). Here U.S. born status along with the young and male domains showed significant differences between the treatment and control groups.

[*] While there is overlap of the variables used in the three metrics, an item might be edited, but not imputed, or imputed and not edited. So the editing flag score and the imputation flag score are measuring different aspects of quality from the same set of variables. The editing flag and critical item flag scores are similar - the critical item flag score is based on a subset of the most important analytic variables from the total group of edited variables. One would expect to find a high correlation between the two scores, but because of critical item retrieval and critical item only complete surveys, we thought it would be useful to calculate and evaluate both scores.

The mean number of Critical variables requiring editing for the Control group was 6.7; the mean number of Critical variables requiring editing from the Brochure incentive group was 11.1 (p=.0475). While there was overall significance, there were no significant subdomains.

Table 1. Quantitative Analysis/Response Rates for Incentive Groups: 2003 SDR Brochure Incentive Experiment

Sampling Characteristics	Control Group - NSF Letter Only			Treatment Group - Brochure			Overall			TREND (Treat. vs. Control)
	Sample Size	Complete Surveys	Response Rate	Sample Size	Complete Surveys	Response Rate	Sample Size	Complete Surveys	Response Rate	
OVERALL	**4,061**	**960**	**24.67%**	**400**	**97**	**24.25%**	**4,461**	**1,054**	**24.64%**	**-0.42%**
SEX GROUPS										
Male	2,854	636	23.37%	279	68	25.09%	3,133	704	23.52%	1.72%
Female	1,207	324	27.75%	121	26	22.31%	1,328	350	27.26%	-5.44%
RACE/ETHNICITY GROUPS										
American Indian	44	14	31.82%	3	1	33.33%	47	15	31.91%	1.52%
Asian	824	217	27.91%	79	18	24.05%	903	235	27.57%	-3.86%
Black	279	86	31.18%	31	7	25.81%	310	93	30.65%	-5.38%
Hispanic	181	70	40.88%	23	10	43.48%	204	80	41.18%	2.59%
Missing	220	15	8.18%	26	3	11.54%	246	18	8.54%	3.36%
Pacific Islander	4	2	50.00%	0	0	-	4	2	50.00%	-
White	2,509	556	23.00%	237	55	23.63%	2,746	611	23.05%	0.63%
CITIZENSHIP AT BIRTH										
Non-U.S. Born	1,261	304	25.61%	127	33	26.77%	1,388	337	25.72%	1.16%
U.S. Born	2,800	656	24.25%	273	61	23.08%	3,073	717	24.15%	-1.17%
PAST RESPONSE										
'01 Cooperated	2,025	738	37.58%	221	77	35.75%	2,246	815	37.40%	-1.83%
'01 Non-response	1,848	165	9.69%	161	13	8.70%	2,009	178	9.61%	-0.99%
New	188	57	32.98%	18	4	22.22%	206	61	32.04%	-10.76%
AGE GROUPS										
<35	385	117	31.69%	41	5	12.20%	426	122	29.81%	-19.49%
35-44	1,185	329	28.69%	108	33	30.56%	1,293	362	28.85%	1.86%
45-54	1,205	297	25.64%	122	29	25.41%	1,327	326	25.62%	-0.23%
55-64	973	168	18.29%	95	22	24.21%	1,068	190	18.82%	5.92%
>64	313	49	16.93%	34	5	14.71%	347	54	16.71%	-2.23%

Note: This table reflects the racial group assigned for sampling purposes. Non-Hispanic multi-race cases were assigned to the racial group that gave them the highest probability of selection. For more detail see the 2003 SDR Sampling Plan.

Table 2: Qualitative Analysis: 2003 SDR Brochure Incentive Experiment

Sampling Characteristics	Control Group - NSF Letter Only		Treatment Group - Brochure		Item Nonresponse (Imputation Score) Comparison			Edit Flag (Variable) Score Comparison		
	Sample Size	Complete Surveys	Sample Size	Complete Surveys	t-value*	p-value	Significant?	t-value*	p-value	Significant?
OVERALL	**4,061**	**960**	**400**	**97**	**-2.26**	**0.0105**	**Yes**	**-2.58**	**0.0112**	**Yes**

SEX GROUPS										
Male	2,854	636	279	68	-2.21	0.0301	Yes	-2.19	0.0316	Yes
Female	1,207	324	121	26	-0.69	0.4950	No	-1.34	0.1925	No
RACE/ETHNICITY GROUPS										
American Indian	44	14	3	1	0.69	0.5027	No	0.66	0.5219	No
Asian	824	217	79	18	-1.34	0.1961	No	-1.01	0.3262	No
Black	279	86	31	7	-0.81	0.4431	No	-1.22	0.2669	No
Hispanic	181	70	23	10	-1.92	0.0817	No	-1.97	0.0770	No
Missing	220	15	26	3	-0.84	0.4729	No	-1.48	0.2726	No
Pacific Islander	4	2	0	0	-	-	-	-	-	-
White	2,509	556	237	55	-0.99	0.3267	No	-1.06	0.2914	No
CITIZENSHIP AT BIRTH										
Non-U.S. Born	1,261	304	127	33	-1.43	0.1609	No	-1.39	0.1735	No
U.S. Born	2,800	656	273	61	-1.67	0.0997	No	-2.10	0.0395	Yes
PAST RESPONSE										
'01 Cooperated	2,025	738	221	77	-1.81	0.0742	No	-1.79	0.0767	No
'01 Non-response	1,848	165	161	13	-1.08	0.2994	No	-1.43	0.1761	No
New	188	57	18	4	-0.51	0.6463	No	-0.94	0.4154	No
AGE GROUPS										
<35	385	117	41	5	4.22	<.0001	Yes	3.74	0.0005	Yes
35-44	1,185	329	108	33	-1.41	0.1671	No	-1.18	0.2451	No
45-54	1,205	297	122	29	-1.51	0.1411	No	-1.58	0.1244	No
55-64	973	168	95	22	-0.79	0.4358	No	-1.53	0.1381	No
>64	313	49	34	5	-0.22	0.8379	No	-0.57	0.5947	No
					* negative values indicate Brochure worse than Control			* negative values indicate Brochure worse than Control		

Table 3 summarizes the Response Rate, Imputation Score, and Edit Flag Score findings.

The Overall row is the most telling. It is here that we clearly see the result that the Brochure incentive had no statistically significant effect on the response rate, but it did have a statistically significant negative effect on all data quality measures for the respondents in the Brochure incentive group.

A look at the statistically significant findings by Sex, Race/Ethnicity, Citizenship Status at Birth, Past Response to the Survey, and Age Group shows few significant differences, and such findings may well be the kind of thing one expects with small sample size domains.

Table 3: Summary of Quantitative and Qualitative Analyses: 2003 SDR Brochure Incentive Experiment

Sampling Characteristics	Response Rate Comparison	Item Nonresponse (Imputation Score) Comparison	Edit Flag (Variable) Score Comparison	Edit Flag (Critical) Score Comparison
	Significant?	Significant?	Significant?	Significant?

OVERALL	No	Yes†	Yes*	Yes*
SEX GROUPS				
Male	No	Yes*	Yes*	No
Female	No	No	No	No
RACE/ETHNICITY GROUPS				
American Indian	No	No	No	No
Asian	No	No	No	No
Black	No	No	No	No
Hispanic	No	No	No	No
Missing	No	No	No	No
Pacific Islander	-	-	-	-
White	No	No	No	No
CITIZENSHIP AT BIRTH				
Non-U.S. Born	No	No	No	No
U.S. Born	No	No	Yes	Yes
PAST RESPONSE				
'01 Cooperated	No	No	No	No
'01 Non-response	No	No	No	No
New	No	No	No	No
AGE GROUPS				
<35	Yes*	Yes*	Yes*	No
35-44	No	No	No	No
45-54	No	No	No	No
55-64	No	No	No	No
>64	No	No	No	No

Finally, we evaluated the different costs of offering the Brochure incentive as compared to the control treatment of simply recontacting the sample members who were reticent to respond to the SDR.

Table 4a describes the cost of the Brochure incentive treatment as compared to the control group baseline. That is to say, the standard costs of processing cases (standard postage, producing mailout materials, and processing returned data) are excluded from this analysis. The only costs measured are the *additional* costs incurred as a result of implementing a particular incentive.

Table 4a: Treatments Costs*: 2003 SDR Incentive Experiment

	Treatment Group	
Costs	**Control**	**Brochure**
Brochure Printing	-	$1,600
Additional Postage	-	$92
TOTAL	$0	$1,692

- These costs exclude standard postage, materials, labor, & management costs, which are assumed to be identical for both the treatment and the control treatment group

Table 4b shows the additional brochure cost of $1,692 distributed by the number of cases worked, that is the number of cases mailed to, and the number of cases completed. As discussed above, this cost calculation

† Brochure group worse than control group.

excludes the standard costs of basic mailing and data processing, which was the same for cases whether they did or did not receive a brochure, and focuses only on the additional cost associated with printing and mailing a brochure.

Table 4b: Additional Cost Per Case: 2003 SDR Incentive Experiment

Treatment	Sample Size	Complete Cases	Cost per Case Worked	Cost per Complete Case
Brochure	400	94	$4.91	$18.00
Control	4,061	960	n/a	n/a

Tables 4a and 4b show that the brochure simply adds cost to the data collection methodology. Our prior quality analysis showed that including the brochure did not yield a statistically significant improvement in response rate between the treatment and control groups. Thus, the increased cost of including the brochure is one more argument against the use of the brochure incentive.

These results are easiest to see as summarized in Table 5.

Table 5. Summary Results: 2003 SDR Brochure Incentive Experiment

Treatment	Cooperation Rate	Item Non-response Mean Score	Edit Flag Mean Score	Additional Cost Per Case Completed
NSF Brochure	24.30%	17.9	10.2	$18.00
Control	24.70%	12.0	6.5	$0.00

The NSRCG Experiment

The NSRCG embedded a brochure experiment into the initial mail contact. Response to the initial mail contact is the least expensive data collection and thus finding methods to increase the response rate at this phase of the survey is of great interest to SRS. The experiment was implemented rather late in the planning process. In doing so, another treatment was inadvertently confounded with the brochure experimental groups; thus we will not discuss this data in detail. While the confounded treatments are not likely to be correlated, making us feel that the results of the brochure experiment from this survey has value since this experiment also finds no statistical differences in response rates between the brochure and control groups.

The NSCG Experiment

The NSCG also embedded a brochure experiment into the initial mail contact as part of a larger experiment. Other aspects of the experiment can be found in Redline, Oliver and Fecso (2004). Identifying potential methods to improve response to the initial mailing was the objective here as well. The experiment related to the brochure split the sample into four random groups. Two versions of appeal letters were crossed with including a brochure or not. Table 6 presents the response rates by a cut-off date at the time of the second mailing for the total and important estimation domains. (Table 7 has the sample sizes.)

Table 6. 2003 NSCG: 1st Mailout Cover Letter Response Rates by Occupation Group

Occupation	Benchmark Letter with Brochure	Benchmark Letter without Brochure	Alternate Appeal Letter with Stylistic Changes with Brochure	Alternate Appeal Letter with Stylistic Changes without Brochure
A Physical Scientists	35.91	33.63	32.98	32.20
B Computer and Information Scientists	24.22	27.84	27.47	27.55
C Life and Social Scientists	37.38	37.11	36.52	33.52
D Engineers	32.25	33.15	33.63	32.51
E Health Occupations	28.43	25.71	29.36	27.85
F S&E related (S&E likelihood group 1)	24.23	25.60	24.84	25.13
G S&E related (S&E likelihood group 2)	30.55	26.45	28.38	31.04
H Teaching	30.87	32.69	32.72	31.66
I Non-S&E Occupations	27.05	25.62	26.97	26.30
Total	***28.85***	***28.63***	***29.44***	***28.89***

Table 7. 2003 NSCG: 1st Mailout Cover Letter Occupation Group Frequencies

	Occupation	Benchmark Letter with Brochure	Benchmark Letter without Brochure	Alternate Appeal Letter with Stylistic Changes with Brochure	Alternate Appeal Letter with Stylistic Changes without Brochure
A	Physical Scientists	596	565	573	559
B	Computer and Information Scientists	1,540	1,552	1,547	1,561
C	Life and Social Scientists	503	512	523	540
D	Engineers	1,445	1,472	1,451	1,455
E	Health Occupations	1,210	1,202	1,209	1,210
F	S&E related (S&E likelihood group 1)	945	961	942	951
G	S&E related (S&E likelihood group 2)	1,031	1,032	1,043	1,034
H	Teaching	2,229	2,227	2,213	2,214
I	Non-S&E Occupations	4,263	4,289	4,242	4,274
	Total	***13,762***	***13,812***	***13,743***	***13,798***

While there are striking differences in response rates across the occupations, within treatments (possibly an indication of a saliency decision), this is not the focus of this research. For our focus on the brochure experiment, overall and within occupations, the brochure and brochure/letter-type interactions had no significant effects.

Summary and Conclusions

The 2003 SDR Brochure Incentive experiment results yield one very interesting conclusion and some indications for directions future research might take.

We conclude from these results that it is not advisable to use a Brochure as a refusal conversion incentive at the end of data collection in the Survey of Doctorate Recipients. There was no statistically significant change in response rates among those receiving the Brochure incentive. For the Brochure incentive treatment group, there were statistically significant decreases on all the measures of data quality; the treatment group showed increases in item nonresponse and in the amount of data cleaning required for acceptable study data. This finding contradicts the commonsense (and anecdotally supported) view that survey participants want to see the results of the study and that those results will motivate their continued participation in the study.

When analyzing the experiment results at a finer level of detail, statistically significant differences were found, but they were limited to three subgroups—men, the U.S.-born, and those under 35 years of age. In each case, the differences in each group are in a single direction. Further experimental research is necessary to shed light on whether there is a causal link between the NSF Brochure and the observed data quality differences.

We also conclude that the use of a brochure as an up-front enticement has no effect on response from our college graduate populations. Given the potential that the brochure might also impact data quality negatively in these groups, use of a brochure is not recommended.

The 2003 SDR Incentive experiment was undertaken on a very particular sample, namely, US Science and Engineering doctorate holders in a longitudinal labor force survey who were in the 29% of the sample that were

nonresponders when the main data collection effort was coming to a close. The college graduate populations are also much smaller and different than the general population. The results clearly aren't generalizable to all populations without further study. However the results lead us to speculate that further research in this area could be fruitful. In particular, it is critical to replicate this experiment with other populations. It is also important to replicate this experiment during other phases of data collection and not just with the most difficult cases at the end of data collection.

Finally, our experience suggests that for studies that employ data editing and cleaning procedures, building in a system of flags to indicate whether a particular variable was changed by particular types of edits/cleaning statements yields a wealth of information that can effectively be analyzed to assess data quality for individual cases. This adds another robust analytic dimension beyond a simple item nonresponse measure.

***Disclaimer**:* This paper reports the results of research undertaken by staff at the National Science Foundation, the U.S. Census Bureau and the National Opinion Research Center (NORC) at the University of Chicago. The views expressed are attributable to the authors and do not necessarily reflect those of NSF, the Census Bureau, or NORC.

Acknowledgements: The authors would like to thank Julia Oliver, John Tsapogas, Kelly Kang, Jeri Mulrow and Mary Frase at NSF for their assistance and contributions. Scott Sederstrom and Steven Pedlow of NORC were also instrumental in developing accurate statistical analyses for this paper. Census Bureau staff at the Jeffersonville Processing Center, especially Fred McKee, as well as John Finamore of DSMD, were instrumental in making the NSCG experiments successful.

References

Church, A.H. (1993). Estimating effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly,* vol. 57 (1), pp. 62-79.

Datta, A.R., Horrigan, M.W., and Walker, J.R. (2001). Evaluation of a Monetary Incentive Payment Experiment in the National Longitudinal Survey of Youth, 1997 Cohort. Presented at the 2001 Federal Committee on Statistical Methodology Conference. Available online at http://www.fcsm.gov/01papers/Horrigan.pdf

Dillman, D.A. (1999). *Mail and Internet Surveys – The Tailored Design Method*, 2nd Ed. New York: Wiley & Sons.

Fecso, R.S. (2001). Experiences with Incentives in a Public Attitudes Survey Using RDD. Presented at the 2001 Federal Committee on Statistical Methodology Conference. Available online at http://www.fcsm.gov/01papers/Fecso.pdf

Fox, R.J., Crask, M.R., and Kim, J. (1988). Mail survey response rate: a meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*, vol. 52, pp. 467-491.

Grigorian, K, Hoffer, T, and Broach, R.J. (2002). 2003 SDR Experiment Summary Plan — Amended. Prepared for NSF by NORC. Issued November 20, 2003.

Grigorian, K, Hoffer, T, and Broach, R.J. (2005). 2003 Survey of Doctorate Recipients Methodology Report. Prepared for NSF by NORC. Issued August 2005.

Gunn W.J., Rhodes I.N. (1981). Physician response rates to a telephone survey: effects of a monetary incentive level. *Public Opinion Quarterly*. vol. 45, pp. 109-115.

Harvey, L. (1987). Factors Affecting Response Rates to Mailed Questionnaires: A Comprehensive Literature Review. *Journal of the Market Research Society*, vol. 29, pp. 341-353.

Hopkins KD, Gullickson AR. (1993). Response rates in survey research: a meta-analysis of the effects of monetary gratuities. *Journal of Experimental Education*. vol. 61:1, pp. 52-62.

Kovac, K.D. and Markesich, J. (2002). Tiered Incentive Payments: Getting the Most Bang for Your Buck. Presented at the Annual Conference of the American Association of Public Opinion Research, May 18, 2002 and published online at http://www.amstat.org/sections/srms/Proceedings/y2002/Files/JSM2002-000130.pdf.

O'Brien, J.E., Levin, K.Y., Hagerty-Heller, T., Greenlees, J.B., and Kirsch, A.D. (2001). Mode and incentive effects on aspects of survey administration and data quality. Available at http://www.amstat.org/sections/srms/Proceedings/y2001/Proceed/00365.pdf.

Redline, C., Oliver, J., Fecso, R. (2004.) The Effect of Cover Letter Appeals and Visual Design on Response Rates in a Government Mail Survey. Presented at the Annual Conference of the American Association of Public Opinion Research, May, 2004 and published in the Proceedings of the Survey Research Methods Section, American Statistical Association.

Singer, E., Groves, R. M. and Corning, A.D. (1999). Differential Incentives: Beliefs about Practices, perceptions of equity and effects on survey participation. *Public Opinion Quarterly,* vol. 63, pp. 251-260.

Tzamourani, P. and Lynn, P. (1999). The effect of monetary incentives on data quality – Results from the British Social Attitudes Survey 1998 experiment. Centre for Research in to Elections and Social Trends. Working Paper Number 73. Available online at: http://www.crest.ox.ac.uk/p73.htm.

THE BROCHURES

NSCG.pdf NSRCG.pdf SEIND04.pdf