Nonresponse Bias Analysis Using Reluctant Respondents in the 2003 National Survey of Recent College Graduates

[1]Donsig Jang, [2]Xiaojing Lin, and [3]Kelly H. Kang
[1, 2] Mathematica Policy Research, Inc., 600 Maryland Avenue, SW, Suite 550, Washington, DC 20024
[3]National Science Foundation, Division of Science Resources Statistics, 4201 Wilson Boulevard, Room 965S, Arlington, VA 22230

Abstract
Reluctant, unmotivated respondents have resulted in increasing data collection costs as survey researchers attempt to maintain traditionally high response rates. This respondent reluctance may be related to the amount of time it takes to complete an interview in large-scale, multi-purpose surveys, such as the National Survey of Recent College Graduates (NSRCG). Recognizing that respondent burden or questionnaire length may contribute to lower response rates, in 2003 the NSRCG offered its nonrespondents monetary incentives about two months before the end of the data collection, following several months of data collection under the standard data collection protocol. In conjunction with the incentive offer, the NSRCG also offered persistent nonrespondents an opportunity to complete a much-abbreviated interview consisting of a few critical items. The late respondents who completed the interviews as the result of the incentive and critical items-only questionnaire offers may provide some insight into the issue of nonresponse bias and the likelihood that such interviewees would have remained survey nonrespondents if these refusal conversion efforts had not been made.

In this paper, we define "reluctant respondents" as those who responded to the survey only after extra efforts were made, beyond the ones initially planned in the standard data collection protocol. Specifically, reluctant respondents in the 2003 NSRCG are those who responded to the regular or shortened questionnaire following the incentive offer. Our conjecture was that the behavior of the reluctant respondents would be more like that of nonrespondents than of respondents to the surveys. This paper describes an investigation of reluctant respondents and the extent to which they are different from regular respondents. We compare different response groups on several key survey estimates. This comparison will expand our understanding of nonresponse bias in the NSRCG, and of the characteristics of nonrespondents themselves, thus providing a basis for changes in the NSRCG weighting system or estimation procedures in the future.

Key Words: Data collection protocol, Response rate, Regular respondent, Weighting adjustment

1. Introduction

Nonresponse is a persistent problem in any meaningful data collection. In particular, surveys in the past three decades have been witnessing continuously declining response rates (de Leeuw and de Heer 2002). With growing concern about privacy, and reluctance to spend personal time in responding to surveys in this information-demanding era, it is expected that response rates will be no better in the foreseeable future and may continue to decline, even with more expensive efforts. Though different factors may be responsible in different surveys, in general two reasons for nonresponse apply to almost all surveys: (1) failure to locate respondents and (2) refusal to participate. Obviously, if we cannot locate the respondents and thus fail to contact them, there is zero probability of getting them to respond to the survey. The other group of nonrespondents consists of those who are contacted and asked to participate in the survey, but refuse to do so. Among many different reasons for refusal, nonrespondents may simply state that they do not want to respond to a survey because they do not have time or interest in it.

The problem for nonresponse bias, however, is that the true bias is not observed. Several attempts are usually made to assess nonresponse bias. Because variables on the sampling frame are available for all sampled cases whether they

responded or not, survey statisticians often compare estimates on frame variables between respondents and nonrespondents to assess any potential bias. If affordable, the survey may even try more intensive data collection efforts such as call-backs or in-person interviews of a subsample of nonrespondents (Elliott et al. 2000). Recently, survey researchers have come to acknowledge the merit and importance to bias assessment of information gathered during the data collection, such as interview date and special dosages to boost response rates including advance letter, reminder letter, incentive offering, and so on (Groves and Heeringa 2006). With ever-decreasing response rates, statistical analysis to assess nonresponse bias and thus determine appropriate weighting adjustments to account for nonrespondents has never been more important.

In this paper, we assess the nonresponse bias indirectly, with the assumption that a certain group of respondents are similar to nonrespondents. Specifically, if a usual data-collection protocol is falling short of getting a certain group of people to respond, extra efforts are often made, including incentive offering or a shortened questionnaire. This paper describes an empirical investigation to assess if extra data collection efforts are worthwhile to compensate for nonresponse bias by persuading at least some reluctant respondents who would otherwise have remained nonrespondents.

Now assume that there is no hard-core group of nonrespondents who absolutely would not respond to the survey, no matter how much effort is made. That is, everyone contacted is assumed to have a non-zero probability to respond to the survey, given the data-collection protocol employed. Such response probability would increase as more efforts are made in terms of time, money, and flexibility. For example, the data collection period might be extended to allow for refusal-conversion efforts. But such an extension of the data collection by a few more weeks or even months would require more resources and thus escalate the costs, and more importantly, delay the reporting schedule. So even with an extended schedule, there is a need to expedite cooperation from nonrespondents. One popular choice is to offer monetary incentive to nonrespondents who have not yet responded to the survey but who would be likely to do so if incentives or some other offers favorable to them are made (REF). The other effort for refusal conversion discussed in this paper is to trim the length of the survey questionnaire, based on the belief that the longer the survey, the less the likelihood of a response. Some surveys have a subset of critical items that must be completed to be a unit response, so in cases of very hard refusals, the surveyor might try to acquire information for those critical items at a minimum. In this paper we call those who would respond to a survey only after extra efforts such as incentives or survey shortening, "reluctant respondents."

Empirical investigation was made for nonresponse bias assessment using reluctant respondents of the 2003 NSRCG data. The behavior of these reluctant respondents appears to be more like that of nonrespondents, however, the current NSRCG weighting system treats these reluctant late respondents the same as early respondents. If differences exist in the characteristics of late and early respondents, it may be possible to develop a special weighting adjustment that accounts for the former.

Section 2 briefly discusses the sample design and data collection procedures for the 2003 NSRCG. Section 3 presents details about the methodology used for the nonresponse bias analysis and provides results from our investigation of nonresponse bias analysis using 2003 NSRCG data. Section 4 provides summaries and recommendations for future surveys.

2. National Survey of Recent College Graduates

The NSRCG was first launched in 1974 and has been conducted every two or three years since then under the sponsorship of the National Science Foundation (NSF). This survey collects education, employment, and demographic information from recent graduates with bachelor's or master's degree in science, engineering, or health (SE&H) fields. This survey has provided an influx to the national data system of scientists and engineers to cover new graduates[1]. It used a two-stage sample design, with a sample from 300 schools at the first stage and a sample of 18,000 graduates from selected schools at the second stage.

[1] NSF maintains the Scientists and Engineers Statistical Data System (SESTAT), which comprises data from three independent surveys: National Survey of College Graduates (NSCG), Survey of Doctorate Recipients (SDR), and NSRCG. For details on SESTAT see http://www.nsf.gov/statistics/sestat/.

The NSRCG used a Computer-Assisted Telephone Interview (CATI) as the primary data-collection mode, with a small mail follow-up in the 1990s and 2001 for nonrespondents who were without telephone numbers. Due to huge cost increases as well as concern about response-rate decrease in telephone surveys, a mixed-mode data collection that included mail and web with CATI follow-up was proposed and implemented in the 2003 NSRCG. However, it was feared that this dramatic change in data-collection protocol might cause differences that would make trend analysis over years difficult. Therefore, in order to be able to trace the source of changes over time, a little over 10 percent of the sample was randomly assigned to a CATI-only group receiving the same data-collection protocol as previously. Zhanyun and Jang (2008) have undertaken an assessment of any effects due to data-collection protocol change.

After five months of standard data-collection protocol with a mixed mode in 2003, the response rate was 45 percent. With this lower-than-expected response, an incentive option was seriously considered at that point to boost the response rate substantially in the next two or three months. To make sure an incentive offer would work, an incentive experiment was conducted by splitting a subset of nonrespondents into "with incentive" and "without incentive" groups. The incentive group showed a significant gain in response rates after one month of the experiment, so the incentive offering was extended to all nonrespondents and the data collection efforts continued for another two months. During this last two-month time period, a small number of hard-core nonrespondents were allowed to respond to only a handful of critical items on the survey. About 1.5 percent of the final respondents responded to such a shortened questionnaire with critical items only. At the close of the 2003 NSRCG data collection, the final response rate was 67 percent.

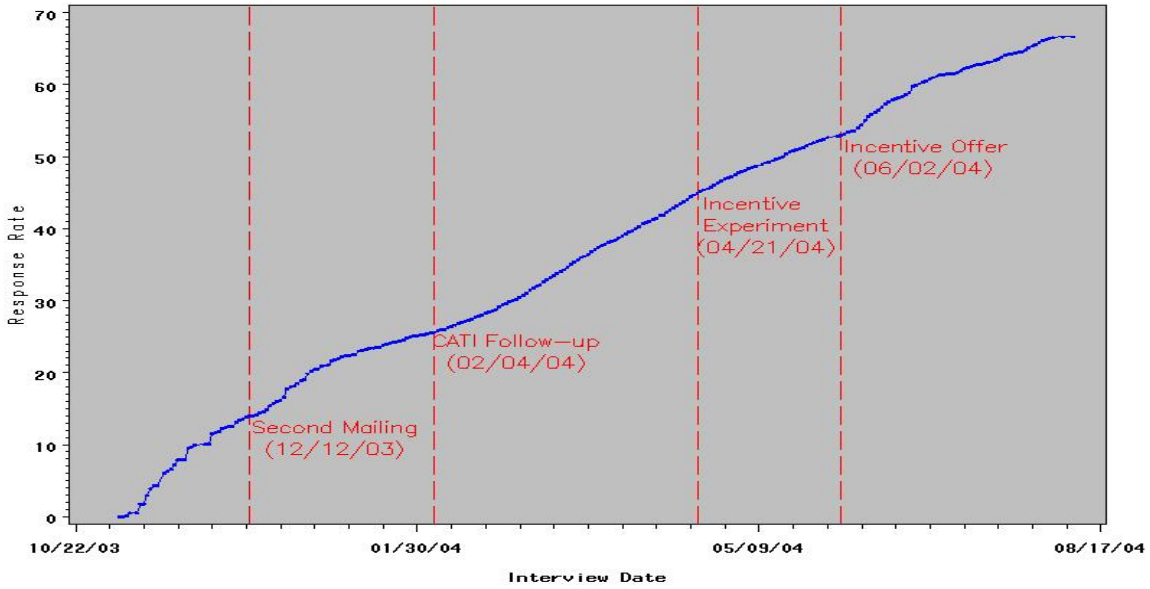

Figure 1: Cumulative Response Rates by Interview Date for the 2003 NSRCG

Figure 1 shows cumulative response rates by interview dates with vertical lines at a few key milestones. An initial rise in the response rate can be noted in the first few weeks. However, it slowed a bit before the second mailing of the questionnaire, which in turn caused another boost in the response rate. A CATI follow-up helped the response rate to rise further, but not as much as expected. The incentive experiment had very little effect on the overall response rate as it applied to only a small subset of nonrespondents. Once the incentive offer was extended to all nonrespondents, it clearly boosted response rates quite a bit. A tiny jump during the extended incentive offering period may indicate extra response rate rise due to critical item-only respondents.

3. Methodology

For nonresponse analysis, all NSRCG sampled cases were classified into five groups based on the following response status:

- Group 1 (RR): Early respondents who responded to the full questionnaire before the refusal conversion incentive offer was made
- Group 2 (RI): Late regular respondents with incentive, who responded to the full questionnaire with the incentive offer
- Group 3 (SI): Late respondents who responded to the shortened questionnaire with critical items only
- Group 4 (LN): Nonrespondents believed to have been contacted
- Group 5 (NN): Not located[2]

Table 1 shows sample counts by response-status category listed above. Note that two sets of numbers are listed in the table: one based on the full sample and the other excluding health majors. With its limited resources, NSF decided to focus on science and engineering majors and thus excluded health major cases from the incentive offering. Therefore, only non-health cases were considered for this analysis.

Table 1: 2003 NSRCG Sample Counts by Response Status

	All Sampled Cases	Excluding Health Degrees*
Regular Respondents (RR)	9,461	8,926
Reluctant Respondents	2,385	2,385
-------Regular Interviews with Incentive (RI)	2,197	2,197
-------Critical Item Interviews with Incentive (SI)	188	188
Nonrespondents (NR)	6,154	5,299
-------Located (LN)	3,080	2,646
-------Not Located (NN)	3,074	2,653
Total	18,000	16,610

*Sampled cases with health degrees were excluded from the monetary incentive offering to focus resources on science and engineering majors. Consequently, health degree cases were excluded from subsequent analyses

Reluctant respondents are defined as those who did not respond to the survey until special treatment such as the monetary incentive offer or critical item-only interview. Such extra efforts indeed helped boost the survey response rate substantially, as presented in Figure 1. This in turn resulted in more cases for analysis. More importantly, those reluctant respondents may provide some insight about the nature of nonrespondents and thus lead to reducing the nonresponse bias.

In differentiating reluctant respondents from regular respondents, we conducted nonresponse bias analysis. Our rationale in doing so is our belief that those reluctant respondents would have been likely to be nonrespondents if the monetary incentive or critical item-only interview had not been offered. With this assumption, we hoped to get some insights on nonresponse bias from those reluctant respondents and help assess bias to account for the bias in the estimation.

Our analyses are twofold: (1) to estimate response propensities by taking into account extra efforts such as the monetary incentive and critical item-only questionnaire offering, and (2) to compare estimates on key survey variables between regular and reluctant respondents.

3.1 Response Propensity Given Data Collection Protocol

Response propensity can be a dynamic attribute, in the sense that it varies depending on the intakes made during the data collection. That is, it is a monotonically nondecreasing function of data collection efforts, including the length of time period, monetary incentive offer, and other options. To understand if such extra efforts would persuade any special

[2] Although some of the noncontact cases may actually have been contacted, we excluded this type of nonrespondent from analysis for the sake of simplicity.

subgroups more than others, we undertook several logistic regression analyses. First, a logistic regression model was run to identify characteristics of regular respondents—those who were more likely to respond to the survey before a serious refusal conversion effort using monetary incentive or shortened interview was made. Second, a similar model was run to identify characteristics of those who were less likely to respond to the full-length questionnaire but who might have changed their mind if a monetary incentive were given. Last, we ran a third logistic regression to identify characteristics of those who were less likely to respond to the full-length questionnaire but might have changed their mind if a short version were available.

A probability to respond to the survey can be calculated in one of the following three ways:

- $p_R(1)$: Treating all respondents the same
- $p_R(2) = p_{RR} + (1 - p_{RR}) p_{RI \cup SI | RR^C}$: Taking into account an incentive offering and the availability of a short version of the questionnaire with critical items combined, where p_{RR} is the probability that a sampled unit would respond to the survey before the incentive offering and $p_{RI \cup SI | RR^C}$ is the probability that a sampled unit would not respond to the survey until extra efforts such as a monetary incentive or shortened interview are offered
- $p_R(3) = p_{RR} + (1 - p_{RR}) p_{RI | RR^C} + (1 - p_{RR})(1 - p_{RI | RR^C}) p_{SI | (RR \cup RI)^C}$: Taking into account incentive offering and shortened questionnaire separately, where p_{RR} is the probability that a sampled unit would respond to the survey before the incentive offering, $p_{RI | RR^C}$ is the probability that a sampled unit would respond to the full questionnaire with the incentive option, and $p_{SI | (RR \cup RI)^C}$ is the probability that a sampled unit would not respond to the full questionnaire but rather to the shortened version of the questionnaire

A separate logistic regression was run to estimate each probability component with sampling variables for covariates such as degree level, degree year, field of major, gender, race/ethnicity, and foreign address.

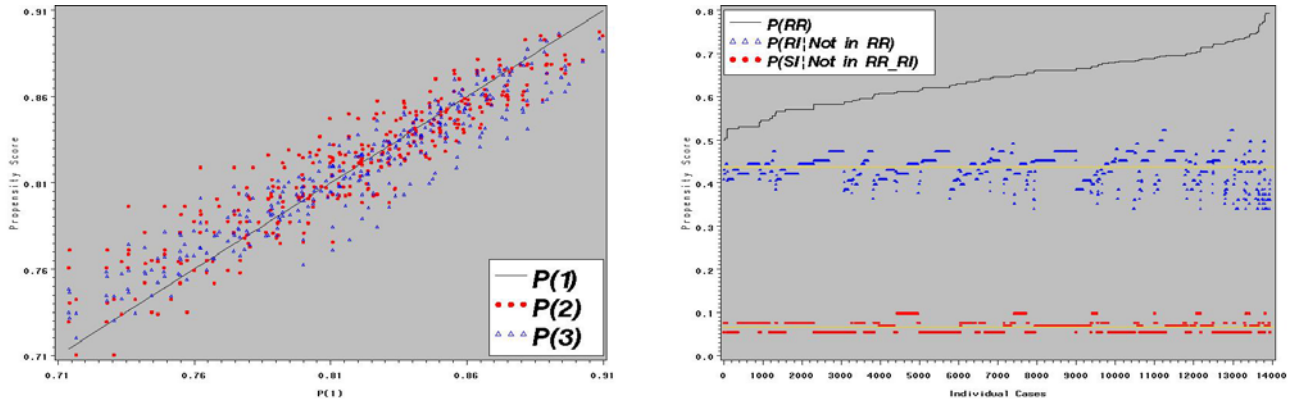

Figure 2: Estimated Response Propensity Distributions

Figure 2 shows two plots: the left shows estimates of the overall response propensity distributions based on each of the three options described above and the right shows response propensity distributions for each of the three individual probability components in option 3.

Though variations are observed, overall response propensity distributions showed strong correlations between the different options. However, predicted values based on option 2 or 3 tend to be higher than ones on option 1 among those who have relatively small propensity values—cases on the left tail. Similarly, cases on the right tail tend to have smaller propensity values based on option 2 or 3 than on option 1. This indicates the importance of utilizing data collection effort information in response-propensity estimation. That is, ignoring the information on extra efforts may

result in underestimation of the response propensity for those who would be more likely to respond after extra efforts are given. The plot on the right shows that estimated response propensity values are almost flat for both extra efforts: incentive offering and shortened questionnaire. This indicates that although extra efforts may help increase response rate, they do not necessarily attract one group more than another.

3.2. Comparison of Survey Estimates

For a population total *Y*, five different estimation options are considered:

- $\hat{P}_{Est1} = \dfrac{\sum_{i \in RR \cup RI \cup SI} w_i^S y_i}{\sum_{i \in RR \cup RI \cup SI} w_i^S}$ where w_i^S is the sampling weight for unit *i* without nonresponse adjustment. This estimator uses all respondents

- $\hat{P}_{Est2} = \dfrac{\sum_{i \in RR} w_i^{A*} y_i}{\sum_{i \in RR} w_i^{A*}}$ where w_i^{A*} is the nonresponse adjusted weight for unit *i*. This estimator uses only regular respondents and treat all reluctant respondents as nonrespondents for weighting adjustments

- $\hat{P}_{Est3} = \dfrac{\sum_{i \in RR \cup RI \cup SI} w_i^A y_i}{\sum_{i \in RR \cup RI \cup SI} w_i^A}$ where w_i^A is the nonresponse adjusted weight for unit *i*. This estimator uses all respondents

- $\hat{P}_{Est4} = \dfrac{\sum_{i \in RR \cup RI} w_i^S y_i + \left(\sum_{i \in SI} w_i^S \right)^{-1} \left(\sum_{i \in SI \cup LN} w_i^S \right) \sum_{i \in SI} w_i^S y_i}{\sum_{i \in RR \cup RI \cup SI \cup LN} w_i^S}$. This is a composite estimator treating only shortened survey respondents as representing all nonrespondents

- $\hat{P}_{Est5} = \dfrac{\sum_{i \in RR} w_i^S y_i + \left(\sum_{i \in RI \cup SI} w_i^S \right)^{-1} \left(\sum_{i \in RI \cup SI \cup LN} w_i^S \right) \sum_{i \in SI} w_i^S y_i}{\sum_{i \in RR \cup RI \cup SI \cup LN} w_i^S}$. This is a composite estimator treating only reluctant respondents as representing all nonrespondents

We calculated estimates for several critical items with each of the five estimation methods described above. For estimates based on Est1 and each of the two reluctant respondents groups (RI and SI), pairwise comparisons were made against estimates based on regular respondents (RR) only. Similarly, estimates based on Est2, Est4, and Est4 were compared to those based on Est3. Relative differences and *p*-values were calculated for each comparison. Table 2 shows estimates for critical items under different estimation options with cells highlighted in different colors based on the magnitudes of absolute relative differences. Specifically, three different colors such as yellow, lavender, and rose indicate absolute relative differences of "between 3 and 5 percentage points", "5 to 10 percentage points", and "larger than 10 percentage points", respectively. The smaller p-values for pairwise t-test, the larger asterisks in cells: "*" for p-values between 0.05 and 0.10, "**" for p-values between 0.01 and 0.05, "***" for p-values between 0.001 and 0.01, and "****" for p-values less than 0.001.

As shown in Table 2, significant differences between regular respondents and reluctant respondents are found. For example, bachelor's degree holders are less likely to respond to the survey unless some incentives are given; percentage estimates for bachelor's degree holders based on reluctant respondents (RI or SI) are significantly higher than those based on regular respondents (RR). Critical item-only respondents have a larger percentage of employed than other response groups. This indicates a potential nonresponse bias if the nonrespondents are more like them. However, due to small sample size for critical item-only respondents, statistical significance does not seem to be strong.

Table 2: Estimates of Critical Items

	With Sampling Weight (Without Nonresponse Adjustment)				With Nonresponse Adjusted		Composite Estimations	
			Reluctant Respondents					
	EST1	RR	RI	SI	EST2	EST3	EST4	EST5
Most recent degree is Bachelor(MRDG)	79.2 ****	77.9	83.2 ****	85.0 ***	79.0 *	79.3	80.3 *	80.0 **
Working for pay during reference week(WRKG)	82.2	82.0	82.8	88.5 *	81.9	82.3	83.5 *	82.4
Looking for work(LOOKWK)	26.9	27.0	26.1	28.4	27.9	27.4	27.2	26.7
Field of study of major for most recent degree:								
Computer and mathematical sciences	15.2	15.0	15.6	19.5	16.1	16.1	16.1	15.3 **
Biological, agricultural and environment sciences	17.7 *	18.3	16.1	10.0 **	16.6 *	16.7	16.1	17.3 *
Physical and related sciences	5.3 ***	5.7	4.2 ***	4.1	4.7	4.7	5.1 *	5.1 ****
Social and related sciences	42.6 ****	40.6	48.8 ****	50.7 *	43.9 **	44.1	44.2	43.8
Engineering	16.6 ****	17.5	13.7 ****	14.9	15.9	15.8	16.3	16.1
Principal job:								
Computer and mathematical scientists	13.4	13.4	13.4	14.2	14.3	14.2	13.6	13.4 **
Biological, agricultural and other life scientists	5.6 **	6.0	4.5 *	2.0 ***	5.5	5.3	4.9 *	5.3
Physical and related scientists	3.5 ***	3.7	2.6 ***	2.1	3.2	3.1	3.2	3.3 ***
Social and related scientists	4.7 ***	5.1	3.3 ***	3.3	5.3 ***	4.7	4.4	4.4 ***
Engineers	13.7 ****	14.7	10.7 ****	9.0 ***	13.2 **	12.9	12.8	13.1
SE related OR Non-SE Occupations	59.2 ****	57.1	65.6 ****	69.4 **	58.6 ***	59.8	61.2	60.5 *

Figure 2 clearly shows that bachelor's degree holders are more likely to respond late and with monetary incentives. However, such potential bias seems to be washed out through nonresponse adjustments, as there seems to be little difference between the two nonresponse adjusted weight-based estimates (Est2 and Est3).

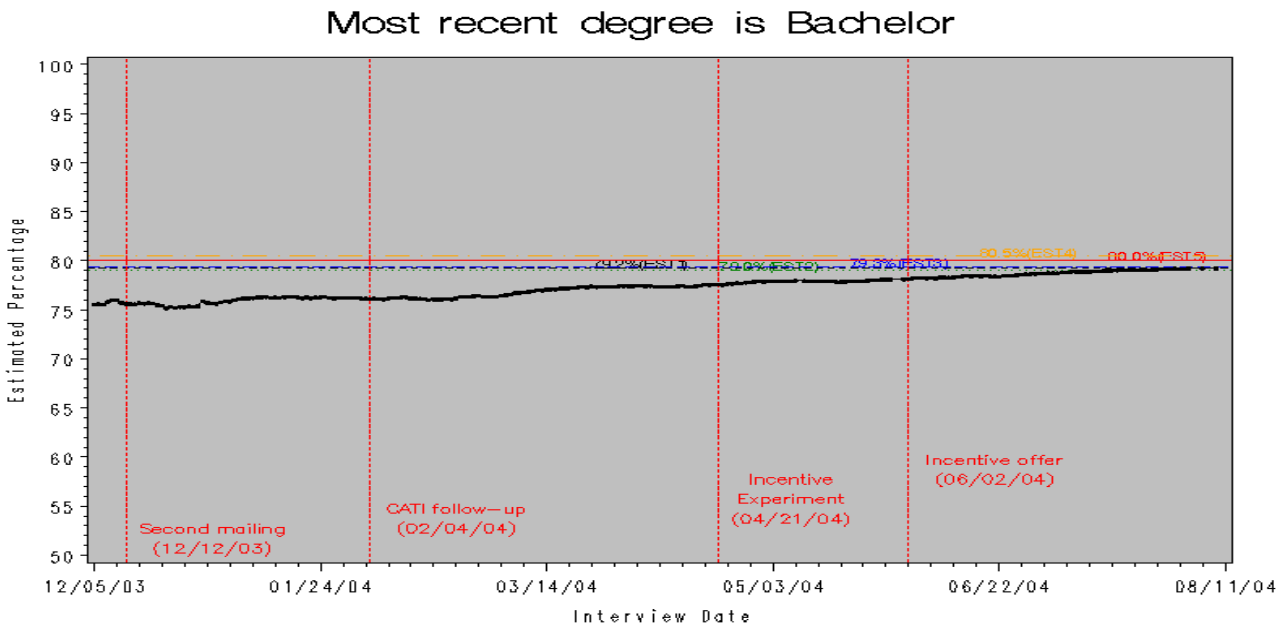

Figure 2: Estimate Trend by Completion Date

4. Summary

In summary, response rates do indeed increase as the incentive and critical item-only interviews are offered, with the rates substantially larger than they would have been without such attempts. Incentive offering worked almost uniformly for all nonrespondents, and slightly better for those who were less likely to respond under the standard data collection protocol. On the other hand, a shortened critical item-only interview helped boost response rate by only a little over 1 percent—and, unlike incentive offering, this option did not specifically gain more responses from whose who were less likely to respond to the full interview. With this option, a serious trade-off occurs for a tiny percent of response rate gain, incurring mass missing data, since the critical item-only respondents leave blank almost all noncritical items.

There are negligible differences between the three options in calculating response propensity values. This indicates that the current weighting approach may be retained, as long as the conventional weighting adjustments are considered.

We considered five different estimation options for population percentage estimates of critical items. These provided empirical evidence that a usual nonresponse weighting adjustment would suffice to reduce potential nonresponse bias for most survey estimates. However, bias could be nontrivial for subpopulation estimates or small percentage estimates. In the future, new weighting adjustments can be considered to compare corresponding estimates with the current survey estimates.

Overall response rate figures may not be informative as the basis for the decision for survey close-out, as low response rates are not necessarily an indication of severe nonresponse bias. Rather, we recommend that survey directors make such a decision in consultation with survey statisticians to focus on domain-specific sizes of respondents, as the sample is often designed to meet analytic objectives for various domains.

Acknowledgements and Disclaimer

Work on this article was supported and funded by the National Science Foundation, contract SRS-0739949. The views expressed here are those of the authors and not necessarily those of the National Science Foundation.

References

De Leeuw, E. and W. de Heer. “Trends in Household Survey Nonresponse: A Longitudinal and International Perspective.” In *Survey Nonresponse*, ed. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little, pp. 41-54. New York: Wiley, 2002.

Elliott, M.R., R.J.A. Little, and S. Lewitzky. “Subsampling Callbacks to Improve Survey Efficiency.” *Journal of American Statistical Association*, Vol. 95, No. 451, 2000, pp. 730-738.

Groves, G.M. and S.G. Heeringa. “Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs.” *Journal of the Royal Statistical Society*, A 169, 2006, pp. 439-457.

Zhao., Z. and D. Jang “Analysis of the Effects of Data Collection Protocol Change in the 2003 NSRCG.” Final Report to National Science Foundation. Princeton, NJ: Mathematica Policy Research, Inc., 2008.