# An Evaluation of Incentive Experiments in a Two-Phase Address-Based Sample Mail Survey

Daifeng Han
Westat

Jill M. Montaquila
Westat

J. Michael Brick
Westat

Address-based sampling (ABS) with a two-phase data collection approach has emerged as a promising alternative to random digit dial (RDD) telephone surveys for studying specific subpopulations in the United States. In 2011, the National Household Education Surveys Program Field Test used a two-phase ABS design with a postal screener to identify households with eligible children and a postal topical questionnaire administered to parents of sampled children to collect measures of interest. Experiments with prepaid cash incentives and special mail delivery methods were applied in both phases. For the screener, sampled addresses were randomly designated to receive either $2 or $5 in the initial mailing. During the topical phase, incentives (ranging from $0 to $20) and delivery methods (First Class Mail or Priority Mail) were assigned randomly but depended on how quickly the household had responded to the screener. The paper first evaluates the effects of incentives on response rates, and then examines incentive levels for attracting the hard-to-reach groups and improving sample composition. The impact of incentive on data collection cost is also examined.

**Keywords:** response rate; prepaid; subpopulation; sample composition; data collection cost

## 1 Introduction

In response to the declining coverage and response rates of landline random digit dialing (RDD) telephone surveys, address-based sampling (ABS) with two-phase postal (or mail) data collection has emerged as a promising alternative for studying specific subpopulations (Han et al., 2010; Brick, Williams, & Montaquila, 2011). Yet little is known about the effectiveness of the various procedures (e.g., use of mail delivery method, nonresponse follow-up mailings, monetary incentives) for administering these surveys. This paper examines the effect of prepaid monetary incentives on response rate, sample composition, and data collection cost in a two-phase ABS mail study – the National Household Education Surveys Program 2011 (NHES: 2011) Field Test. Sponsored by the United States (US) National Center for Education Statistics (NCES), NHES is a set of periodic education surveys that were conducted as landline RDD surveys until 2007. Like many other landline surveys in the US, NHES experienced a decline in coverage rates and response rates (see Figure 1). Although dual-frame cell phone and landline surveys address the coverage issues, cell phone response rates in the US are lower than those of the landline component (AAPOR Cell Phone Task Force 2010). These

dual frame surveys also face some challenging statistical issues (Brick et al. 2011).

In 2009, NHES tested a mail data collection approach using a two-phase ABS design as an alternative to the RDD methodology. Brick, Williams, and Montaquila (2011) found this alternative was feasible. This article is based on the "NHES: 2011 Field Test", a very large-scale methodological study designed using lessons learned from the 2009 Pilot Study.

### 1.1 The NHES: 2011 Field Test

The overall objective of the "NHES: 2011 Field Test" was to evaluate various experimental conditions for data collection and identify economical alternatives for not only achieving high response rates, but also limiting potential nonresponse bias (since response rate is not necessarily a good indicator of nonresponse bias). The Field Test targeted two subgroups; school-aged children (those in kindergarten through twelfth grade) were eligible for the Parent and Family Involvement in Education (PFI) questionnaire and children ages six years or younger who had not yet started kindergarten were eligible for the Early Childhood Program Participation (ECPP) questionnaire. The need to screen for households with eligible children and to sample from the list of enumerated children dictated a two-phase design. Mail was used for the first phase (screener) survey because mailing address is the only contact information available for all households in the sampling frame; see Iannacchione (2011) for more details on the sampling frame. For the second phase

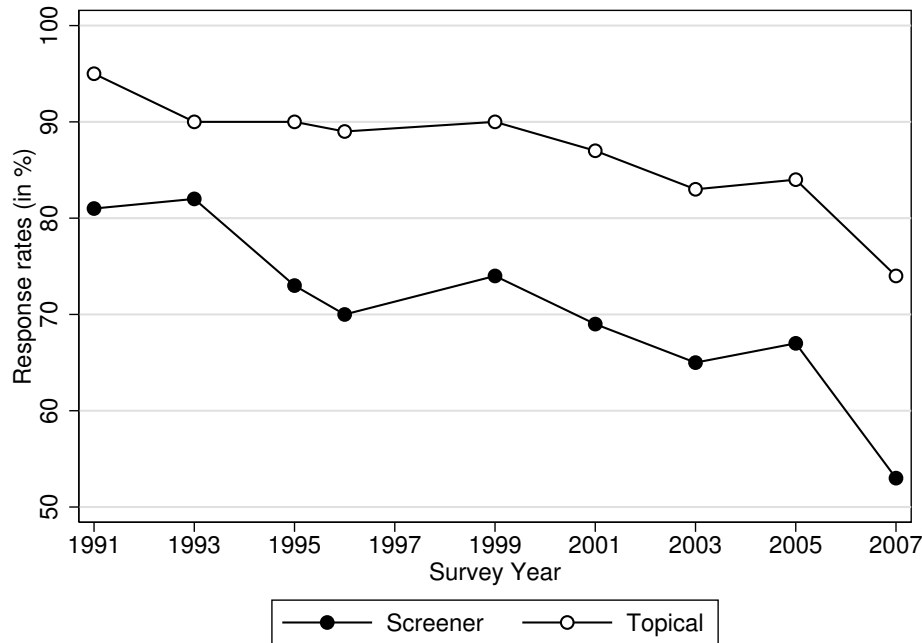Contact information: Daifeng Han, Westat, Rockville, MD, United States, e-mail: DaifengHan@westat.com

*Figure 1.* NHES response rates over time

Source: National Household Education Surveys Program, 1991-2007. Note: Topical response rates shown here are for surveys of children only. (The NHES program has also included topical surveys of adults.)

(topical) survey, we chose the mail mode based on the results of the 2009 Pilot Study.

The screener survey was used to identify households with eligible children. Items collected in the screener instrument included the age, gender, school enrollment, and grade for each person of age 20 or younger living at the sampled address. Any first-phase responding household with at least one child eligible for either the PFI or the ECPP study was randomly sampled for a topical survey by Westat (the data collection organization) using the reported screener enumeration information to ensure a probability sample. One child was sampled per household, and the topical questionnaire was sent to the parents/guardians of the sampled child to collect data on the care and education of the sampled child. Information for the sampled child (i.e., age, gender, school enrollment, and grade that was collected in the screener) was printed on the topical questionnaire, so the parents/guardians would know for which child they should be filling in the topical survey.

Based on the total design method (Dillman et al. 2009), we administered up to three questionnaire mailings in each phase of data collection. For each phase, approximately three weeks after the initial mailing, nonrespondents (excluding postmaster returns) were mailed a second packet. Then the remaining nonrespondents were mailed a final packet about six weeks after the initial mailing. A set of experiments was included to examine the effects on response rate and nonresponse bias (Montaquila et al. 2012), and the house-

holds that were treated were not aware that they were part of an experiment. The experiments reported here varied the amount of the prepaid cash incentive in the initial mailing for both phases, the mail delivery method (the US Postal Service's First Class or Priority Mail delivery methods), and the amount of incentive in the final topical survey nonresponse follow-up mailing. We describe the incentive treatments in greater detail in Section 2.

### 1.2 Research Questions

Two-phase mail data collection is an innovative approach (Brick, Williams, & Montaquila, 2011). The incentive literature that is most relevant to this approach covers single-phase mail surveys, telephone surveys targeting specific subpopulations, and longitudinal surveys. Here we briefly review the existing literature and lay out the questions we aimed to answer through this research.

Although monetary incentives have been used in many surveys, the theory for explaining the effectiveness of incentives has not been fully established. The alternative frameworks include social exchange theory (Dillman 1978), the norm of reciprocity (Gouldner 1960), and economic exchange theory (Biner & Kidd 1994). Groves et al. (2000) proposed a more comprehensive model that viewed incentives as an inducement to compensate for the absence of factors that might otherwise stimulate cooperation (e.g., topic salience, a sense of civic obligation).

Studies have demonstrated that token monetary incentives increase response rate in single-phase mail surveys (Lesser et al. 2001; Church 1993). Prepaid incentives are almost always more effective than promised incentives that are contingent on response (e.g., Petrolia & Bhattacharjee 2009; Church 1993). However, findings are mixed on how large an incentive should be and whether the effects of incentives are linear (e.g., Yu and Cooper 1983; Gelman et al. 2002; Cantor et al. 2008). Prepaid incentives can either be provided to all units in the initial attempt, or only to those that initially failed to cooperate (e.g., Brick et al. 2005; Stoop et al. 2010). Some researchers recommend using refusal conversion payments in combination with sampling nonrespondents as a more cost-effective approach (Brick et al. 2005). The longer-term impact of incentives has also been studied in face-to-face and telephone longitudinal surveys. There is no evidence that an earlier wave incentive would condition the respondents to expect incentives in later waves (e.g., Creighton et al. 2007; Singer et al. 1998). This finding is relevant to the two-phase design of the "NHES: 2011 Field Test".

While prepaid incentives generally increase response rates, the increase does not always decrease nonresponse bias in survey estimates. The linkage between nonresponse rates and nonresponse bias arises from the presence of a correlation between response propensity and the survey variables of interest (Groves et al. 2006).

Offering an incentive may affect nonresponse bias by altering the composition of the respondents. For example, Griffin et al. (2011) and Lesser et al. (2001) showed that incentives decreased potential nonresponse bias by increasing the likelihood of participation by males (who are generally less likely to respond to mail surveys than females). Some studies reported incentives increased response from those less interested in or less familiar with the survey topic (e.g., Groves et al. 2006; Petrolia & Bhattacharjee 2009). Several studies demonstrated that monetary incentives were effective in recruiting low-income, low-education, or minority respondents (Martin et al. 2001; Singer & Kulka 2002; Petrolia & Bhattacharjee 2009). On the other hand, Juster & Suzman (1995) found that offering $100 per individual and $200 per couple in refusal conversion brought into the sample people with higher incomes and more net worth than those who had never refused or had been converted with other efforts. Singer (2002) reviewed the "intended" and "unintended" consequences of using incentives and concluded that while monetary incentives are generally effective, less money is required to recruit and retain low-income groups than those whose income is higher.

Our research was motivated by the concern that findings from incentive experiments in single-phase cross-sectional surveys would not hold in the two-phase setting. For example, would the positive effect of a screener incentive carry over to the topical phase? In the RDD context (based on the NHES: 2003), Brick et al. (2005) concluded that if there was any effect on cooperation rates at the topical interview level that had been carried over from giving incentives at the screening interview level, then the effect was relatively

small. This finding is relevant, but not directly applicable to our mail survey context. Similarly, experiences from longitudinal surveys are relevant, but in our two-phase mail design the time lag between the two questionnaires is much shorter and no interviewers are present to establish rapport as in the face-to-face or telephone data collection modes.

Our first challenge was to identify combinations of screener and topical incentives that result in high overall response rates across both phases. When surveying a specific subpopulation from a general population frame, survey researchers face two situations. If no reliable benchmark is available for the target subpopulation, it is important to elicit roughly equal response rates from all the units in the sample to estimate the prevalence of the specific subpopulation. In the "NHES: 2011 Field Test", however, a reliable benchmark exists for the target subpopulation (households with children) so our primary goal was to identify approaches that maximize responses from the specific subpopulation. Responses from households without children did reduce data collection costs associated with screener nonresponse follow-up mailings. Thus, we also use the percentage of screened households with a child as a measure of the quality of the survey. Since response rates are not good indicators of nonresponse bias (Groves & Peytcheva 2008), we aimed to identify methods that improve the similarity of the respondents to the target population. Household surveys, especially when conducted through mail data collection mode, tend to underrepresent groups with lower socioeconomic status (SES). We examine the base-weighted distribution of the respondents and evaluate how close this distribution is to the distribution of the target population (from external benchmarks) under different incentive treatments. In addition, the incentives and mailing treatments affect the cost of data collection. Therefore, we study whether there are incentive levels that attain a high response rate and low potential nonresponse bias while containing the data collection cost in the two-phase mail design.

Finally, a key to persuading people to cooperate is to improve the experience of participation and reduce response burden. Monetary incentives are only one tool for increasing the motivation for survey participation. It is important to consider all the contributing factors and identify the best overall approach within the available resources. The "NHES: 2011 Field Test" included other experiments such as screener questionnaire versions and mail delivery methods, so we explore interaction effects between these factors and the incentive treatments.

The remainder of this paper is organized as follows: Section 2 describes the incentive, questionnaire version, and mail delivery method treatments in greater detail. Section 3 evaluates the effectiveness of screener and topical incentives in increasing responses from all units in the sample and from the target subpopulation. The impacts of incentives on sample composition and data collection cost are presented in Section 4. We summarize and discuss the findings in Section 5.

## 2 Study Design and Incentive Treatments

The "NHES: 2011 Field Test" included a national sample, along with a supplemental or targeted sample of addresses that were more likely to contain Spanish speakers. In this paper we report the findings only from the national sample – a simple random sample of addresses from the ABS frame. This sample included a "pilot" group (a random subsample) that, by design, received a particular combination of treatments and was not part of the incentive experiments, so those cases were excluded from the analysis. The total sample size for this study was 36,260. This section describes the incentive, questionnaire version, and mail delivery treatments. All the other experiment factors are described in Montaquila et al. (2012) and crossed with the incentive treatments. All response rates given here are AAPOR RR4 rates (AAPOR 2011).

### 2.1 Screener Treatments

All the sampled addresses were mailed an initial screener packet containing a screener questionnaire, an informational letter, a business return envelope and a token monetary incentive; the sampled addresses were randomly designated to receive either $2 or $5. Since the positive effect of prepaid incentives in single-phase surveys has been well established in the literature, we did not include a group with no incentive treatment in the screener phase. A thank-you/reminder postcard was mailed to all the addresses. Nonrespondents to the initial mailing were mailed a second screener packet, and the remaining nonrespondents were mailed a final screener packet. The initial mailing and first nonresponse follow-up mailing were sent through First Class Mail. The second nonresponse follow-up mailing was sent using either FedEx delivery or Priority Mail and the cases were randomly assigned to one of the two treatments. No monetary incentive was provided during the two follow-up mailings.

Two screener questionnaire versions were tested. The Screen-out version was a very short questionnaire that focused solely on identifying children. The Engaging version was longer and included items aiming to interest target respondents. If no children were present in the sampled address, the household needed to answer only one item to complete the Screen-out version. In the Engaging version, the household was asked to respond to approximately 25 items about education and the household before reaching the items about the presence of children in the household. The sampled addresses were randomly assigned to receive either Screen-out or Engaging questionnaire during the initial mailing. During the screener follow-up mailings, half of the households were sent the same version as in the initial mailing, and the other half were sent the different version (i.e., Screen-out switched to Engaging, or Engaging switched to Screen-out). The rationale behind the questionnaire version experiment is explained in Williams et al. (2012).

The screener questionnaire mailing procedure and incentive treatments are shown in Figure 2. We define two terms based on screener response time. If a household responded to the initial mailing or first follow-up mailing, then it is considered a Screener-Early case. If a household responded to the second (final) screener follow-up mailing (which was sent six weeks after the initial mailing), then it is considered a Screener-Late case.

### 2.2 Topical Treatments

One child was sampled from each screener responding household with either ECPP- or PFI-eligible children. The topical incentive and mail delivery method treatments were assigned randomly but depended on how quickly the household had responded to the screener questionnaire – Screener-Early and Screener-Late cases were treated differently.

Figure 3 shows the topical incentive and mail delivery method treatments. The Screener-Early respondents were randomly assigned to six groups for the initial topical mailing – half of the cases were sent Priority Mail with no incentive; the remaining half were split into five groups to receive First Class Mail with either $0, $5, $10, $15, or $20. In contrast, the Screener-Late households had already received either a FedEx or Priority Mail mailing for the final screener nonresponse follow-up, so we did not use special mail delivery for these cases during the initial topical mailing. Instead, we assigned the Screener-Late cases randomly to five levels of incentive treatments (i.e., First Class Mail with $0, $5, $10, $15, or $20). Here treatment groups with no incentive (designated for either First Class or Priority Mail) were included because a connection had already been established with the target respondents during the screener phase, and we wanted to assess the effect of this earlier incentive in the topical phase.

The first topical nonresponse follow-up mailing was sent to all topical nonrespondents through First Class Mail with no incentive. The treatment for the second follow-up mailing depended on the delivery method used in the initial topical mailing. Cases that were sent the initial mailing via Priority Mail received no monetary incentive then, so for the second follow-up mailing they were randomly assigned to receive either $5 or $15 (through First Class Mail) for testing the effectiveness of a higher monetary incentive for nonresponse follow-up at this late stage. On the other hand, cases that were sent the initial topical mailing via First Class Mail were already subject to incentive treatments, so for these cases the second nonresponse follow-up mailing was sent through Priority Mail with no incentive.

In a two-phase mail survey with the sequence of mailings used in the "NHES: 2011 Field Test", a household could receive up to six questionnaire packets, so we varied the incentive and delivery method treatments from one mailing to another to give the target respondents a fresh stimulus with each mailing. For example, one hypothesis was that if the delivery method for the initial topical mailing was different from the screener mailing the household responded to, then the household would be more likely to distinguish the envelope and more likely to respond. The Screener-Early households had responded to First Class Mail packets during the screener, so half of the cases were sent Priority Mail during
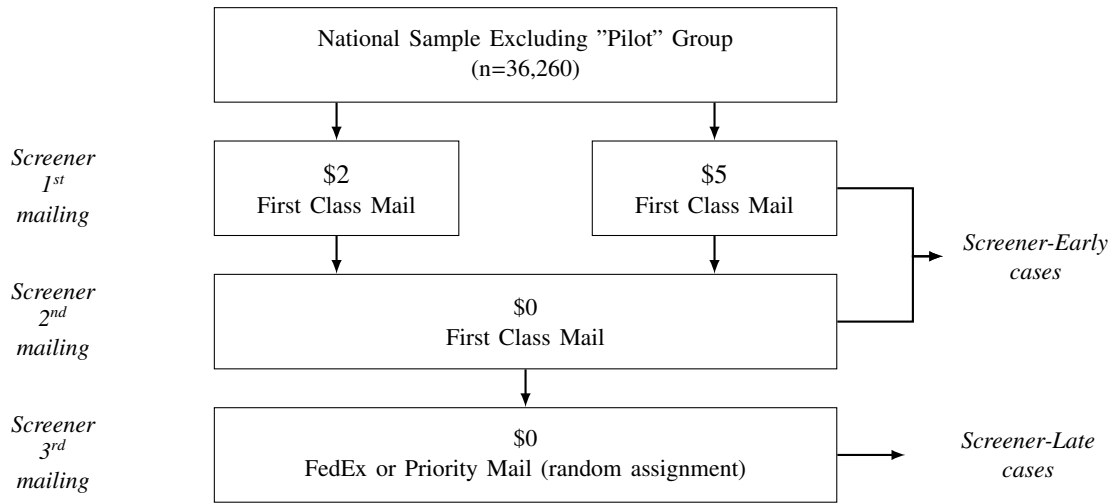
*Figure 2.* Screener incentive and delivery method treatments.

[a] Screener-Early cases are households that responded to the initial or first follow-up mailing.
[b] Screener-Late cases are households that responded to the second screener follow-up mailing.



*Figure 3.* Topical incentive and delivery method treatments

[a] Screener-Early cases are households that responded to the initial or first follow-up mailing.
[b] Screener-Late cases are households that responded to the second screener follow-up mailing.
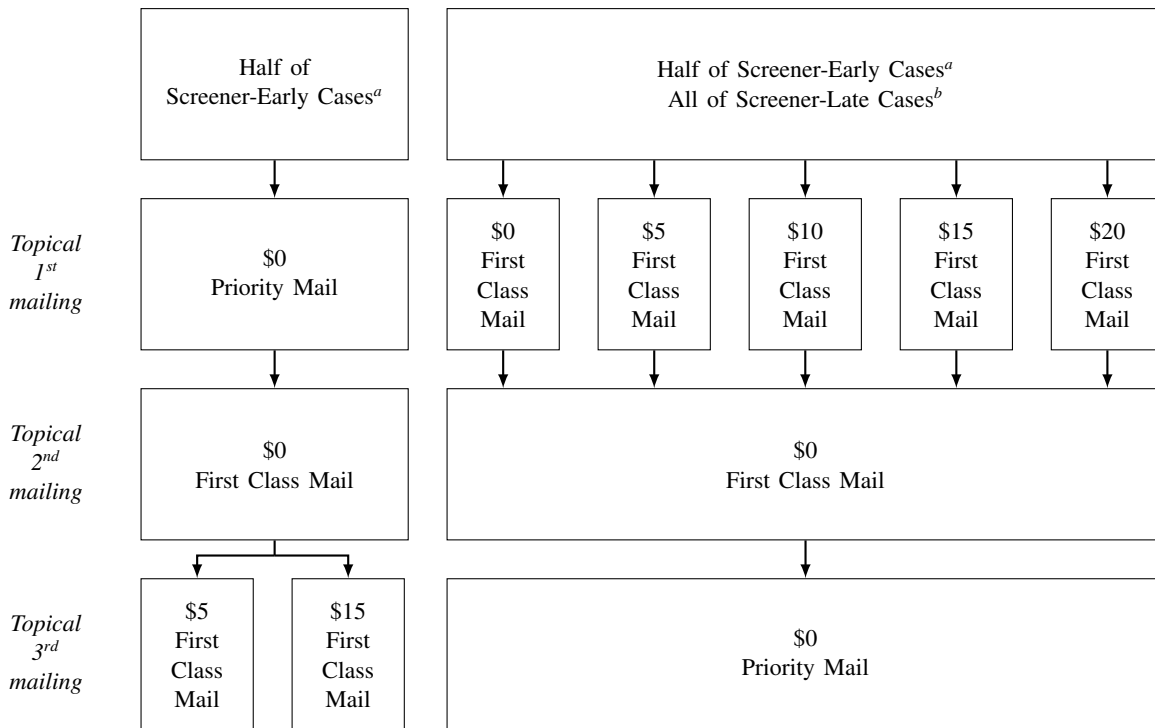
the initial topical mailing. The Screener-Late cases had been exposed to FedEx or Priority Mail during the screener, so their initial topical mailing was sent using First Class Mail. It is worth noting that the envelope featured the US Department of Education sponsorship seal whenever First Class Mail was used, but the FedEx or Priority Mail envelope was the standard envelope used by the delivery service and did not show any sponsorship information.

## 3 Effects on Response Rates

One of the key measures of the effectiveness of the various treatments is the response rate, so this section examines screener response rates and topical response rates (conditioning on screener response) by experimental treatments. Since the "NHES: 2011 Field Test" targeted a specific subpopulation, we also report the eligibility rates from the screener phase – the proportion of screener responding households with eligible children. All the analyses and tabulations are weighted to account for differential probabilities of selection.

### 3.1 Screener Incentives

Table 1 summarizes the effects of screener incentives on screener response rates, eligibility rates, and overall sample yield across both phases. The standard error of the estimate is given in parentheses following the estimate. For the screener phase, we present initial response and eligibility rates in addition to final response and eligibility rates. The initial rates reflect the success of the initial screener mailing, during which either $2 or $5 was offered. The screener response rate was 70.9% for the $5 group and 66.5% for the $2 group. The 95% confidence interval for the 4.4% difference is (3.4%, 5.4%). Additionally, the difference between the initial rates for the two groups indicates that the higher incentive in the initial mailing saved some cost associated with nonresponse follow-up mailings.

A second quality measure is the screener eligibility rate – the proportion of completed screeners in which the household indicated the presence of an eligible child. The screener eligibility rates for the $2 and $5 treatment groups were comparable. The findings about screener response rates and eligibility rates hold for all types of addresses (i.e., city style, rural, highway, and P.O. Box addresses). In a two-phase study, the screening survey serves to identify members of the target population, but the data needed for analysis comes from the topical responses from the target population; thus, topical response rate is also important. The topical response rates (conditioning on screener response) were similar (not statistically significantly different) for the $2 treatment (73.9%) and the $5 treatment (71.9%).

A more comprehensive measure of the effectiveness of the screener treatments across both phases of data collection is the ratio of the number of sampled addresses (excluding undeliverable addresses) to the number of topical completes. A lower ratio implies that we need to sample fewer cases to obtain a topical complete. The ratios are similar 6.4 for the $2 screener incentive group and 6.2 for the $5 screener incentive. These ratios show that the higher screener response rate

for the $5 treatment does not result in higher yield of topical completes because of losses in the percent eligible for the topical and the conditional topical response rates. This finding is consistent with that reported by Brick et al. (2005), who found that the positive effect of higher screener incentive was not carried over to the topical phase in the NHES: 2003 landline RDD administration.

As described earlier, two screener questionnaire versions were used – a very short questionnaire designed to limit respondent burden and a longer questionnaire with education-related questions designed to engage the target population. We found no interaction effect between screener incentives and questionnaire versions using the ratios of sampled addresses to completed topical surveys described above.

### 3.2 Topical Incentives and Mail Delivery Methods

One measure of the effectiveness of the topical incentive treatments is the topical response rate (conditioning on screener response). We conducted two comparisons across various treatment groups using this measure. The first comparison was based on the cases that were sent the initial topical mailing via First Class Mail. For this analysis, the Screener-Early cases that were sent the initial topical mailing via Priority Mail were excluded, and the remaining Screener-Early cases were re-weighted to the distribution of Screener-Early versus Screener-Late cases among all the screener respondents. We refer to this grouping and re-weighting as the "five-group comparison", and it is used in several analyses and tabulations. This grouping was used to compare the effectiveness of the different topical incentive levels regardless of the screener response time.

As shown in Figure 4, the conditional topical response rates for the treatments of $5 or higher were statistically and substantively higher than the rate for cases in the no incentive treatment group. The differences between the $5 treatment and the higher incentive levels ($10, $15, or $20) were negligible. Trussell and Lavrakas (2004) also found that the effect of monetary incentives is nonlinear, with the increase in response rates decreasing as the incentive amount increases.

The slopes in Figure 4 indicate that the marginal returns of additional mailings are approximately equivalent across different incentive levels. There was a steady increase in response rate with each follow-up mailing. It is worth noting that the effects of the second follow-up mailing were due in part to Priority Mail, and a different pattern might hold if First Class Mail had been used. Figure 4 also clearly shows the importance of repeated mailings – the final response rate for the no incentive treatment is higher than the initial mailing response rates for all of treatments with incentives.

For the second comparison, the topical cases are divided into eleven groups based on screener response times; Table 2 shows the results for the Screener-Early and Screener-Late cases in separate columns. Three patterns are apparent. First, screener response time was a good indicator of topical response propensity. Regardless of the topical incentive amount, topical response rates are consistently higher for the Screener-Early cases than for the Screener-Late cases.

*Table 1* Screener incentive treatment effects

| | Screener incentive treatment | |
|---|---|---|
| | $2 n=18,130 | $5 n=18,130 |
| **Screener phase** | | |
| Final response rate* | 66.5% | 70.9% |
| | (0.33%) | (0.36%) |
| Final eligibility rate | 32.3% | 32.4% |
| | (0.46%) | (0.44%) |
| Initial response rate* | 36.3% | 42.8% |
| | (0.39%) | (0.43%) |
| Initial eligibility rate | 28.2% | 28.8% |
| | (0.62%) | (0.58%) |
| **Topical phase** | | |
| Conditional response rate[a] | 73.9% | 71.9% |
| | (0.86%) | (0.98%) |
| **Both phases** | | |
| Number of sampled addresses per topical complete [b] | 6.4 | 6.2 |
| | (0.12) | (0.13) |

Source: National Household Education Survey Field Test, 2011.
Notes: Standard errors are shown in parentheses. * p-value < 0.05. [a] Topical response rates are calculated at the child level, conditional on screener response. [b] The ratios are calculated after accounting for undeliverable addresses.
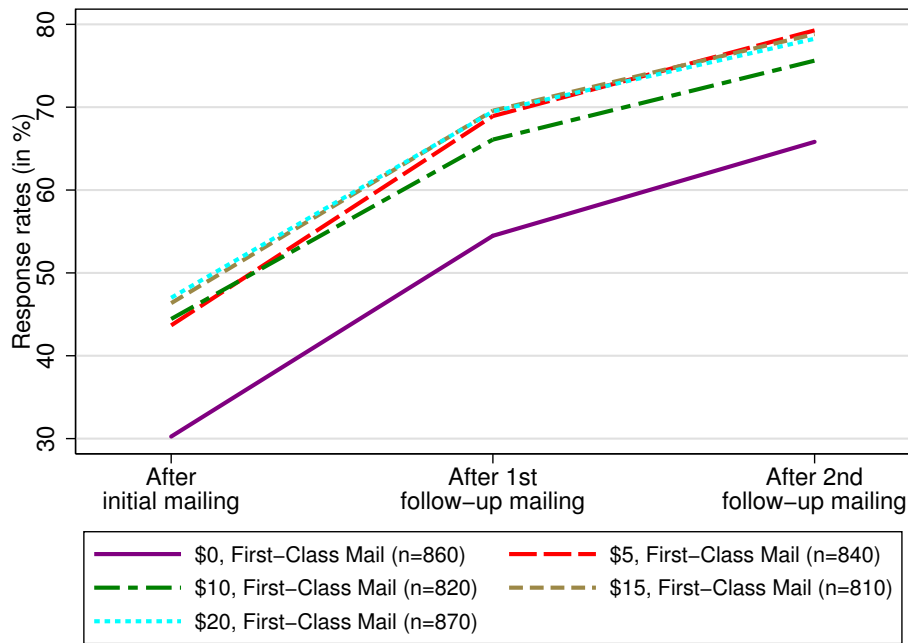


*Figure 4*. Effects of topical incentives on response rates: Five group comparison.

Note: The Screener-Early cases that were sent the initial topical mailing via Priority Mail are excluded from the analysis, and the remaining Screener-Early cases are re-weighted to the distribution of Screener-Early versus Screener-Late cases among all the screener respondents. For all the cases in this analysis, the first and second topical follow-up mailings were sent with no incentive, via First Class Mail and Priority Mail, respectively.

Second, the findings about different incentive levels from the five-group comparison hold for both the Screener-Early and the Screener-Late cases – $5 was effective relative to no incentive, but offering higher incentives did not result in further increases in the response rate. For the Screener-Late cases, the topical response rates associated with $10, $15, and $20 were nominally higher than that of the $5 treatment, but the differences are not statistically significant.

Third, for the Screener-Early cases that were not offered monetary incentives during the initial topical mailing, the difference in the response rates between the Priority Mail treatment (73.6%) and the First Class Mail treatment (72.0%) was small. This was true despite the fact that the Priority Mail cases were offered either $5 or $15 during the second nonresponse follow-up mailing. One possible explanation is that the official government envelope used for the First Class mailing might be about as effective as the Priority Mail envelope for the Screener-Early cases. The postage for the Priority Mail is approximately $5, so using the money as a prepaid cash incentive rather than for Priority Mail postage was the more effective approach to improve response.

Our final response rate analysis focuses on the effects of incentives used for nonresponse follow-up. The cases used for this analysis are those that were sent the initial mailing through Priority Mail with no incentive and did not respond to the first two topical mailings; thus it is restricted to Screener-Early cases. A second nonresponse follow-up mailing was sent to these cases via First Class mail and the cases were randomly assigned to receive a prepaid cash incentive of either $5 or $15. The question was whether the higher incentive level would be more effective in converting the nonrespondents at this late stage. We calculated the conditional response rate to the second nonresponse follow-up mailing by incentive treatment as a measure of effectiveness. The 6.6% difference between the $5 treatment (29.5%) and the $15 treatment (36.1%) has a 95% confidence interval of (-0.7%, 13.9%). We cannot evaluate the effectiveness of the $5 incentive in the second nonresponse follow-up relative to no incentive, because the experimental design did not include a condition with no incentive at this stage. It is also worth noting that only the Screener-Early cases were subject to incentive treatments in the final nonresponse follow-up, so the findings may not apply to all the screener respondents.

For all the topical comparisons described above, we repeated the analyses for the ECPP and PFI children (younger and older children, respectively) separately and found similar results for both subgroups. We also tested the interaction effects between the screener and topical incentives but did not find any significant interactions.

## 4 Effects on Sample Composition and Data Collection Cost

While overall response is important, incentives may be particularly effective in gaining cooperation from low response propensity persons and thus reduce nonresponse bias. To study these effects, we computed estimates of various characteristics by incentive treatments using the unadjusted child sampling weights. The groups included in this analysis and the weighting method are the same as in the five-group comparison presented in Section 3.2, except that only the topical respondents were used to generate the estimates (because the data were only available for respondents). We estimated ten demographic and SES characteristics for each of the five incentive groups. All the estimates were proportions. The variables were coded with value 1 signifying minority or lower SES, and value 0 otherwise. Thus, higher estimates indicated relatively higher response from minority and/or lower SES groups. Benchmark estimates of each of these characteristics were obtained from the American Community Survey (ACS) and the National Health Interview Survey (NHIS). We then calculated the ratio of the NHES estimate to the benchmark estimate for each of the ten variables; this ratio is a standardized measure that indicates the effectiveness of the NHES survey in gaining the participation of minority and/or low SES groups.

Figure 5 shows these ratios by topical incentive level. We combined the $10, $15, and $20 treatments due to the similarity of the estimates. The conclusion is that compared to the no incentive treatment or the higher incentive levels, the $5 topical treatment estimates are generally closer to the benchmarks (a ratio of 1.0). In particular, the $5 topical incentive did consistently better in obtaining response from minority and/or low SES groups. This is a favorable result because mail surveys often get lower response rates from lower SES groups. These results are consistent with the theory posited by Singer (2002), who noted that the threshold for recruiting lower-income groups is lower than the threshold for those whose income is higher. As a result, when higher incentives ($10, $15, and $15) were offered, these incentives were associated wtih a higher percentage of the higher SES people responding. Dillman, Smyth, & Christian (2009) also provide an explanation that higher incentives invoke people to consider this an economic transaction, while lower incentives cause people to feel obliged to return positive behavior as posited by social exchange theory. If the incentive is large, it might be viewed as a bribe or an undue pressure to comply; under those circumstances, compliance is inhibited (Groves et al. 1992). The higher incentive levels might be more likely to create this pressure for the lower SES groups.

These estimates also clearly demonstrate a non-linear relationship between the incentive amount and improvement in sample composition. That is, further increase in incentive levels does not result in further decrease in potential nonresponse bias as indicated by these characteristics. For the specific subpopulation we were interested in, it appears likely that $5 was the optimal topical treatment level. The $5 topical treatment generally seemed to decrease the potential nonresponse bias compared to both the $0 treatment and higher incentive treatments, although we were not able to directly compute estimates of the extent of bias reduction.

A similar evaluation was conducted for the screener incentive treatments. We did not find any significant difference between the two screener incentive levels ($2 versus $5) for any of the ten variables, nor did we see any interaction effects

*Table 2*  Topical conditional response rates, by topical incentive treatment and screener response time

| Initial topical mailing treatment | Screener-Early cases | | Screener-Late cases | |
|---|---|---|---|---|
| | Sample Size | Topical conditional response rate | Sample Size | Topical conditional response rate |
| Priority Mail, $0[a] | 2530 | 76.6%(1.2%) | 0 | *NA* |
| First Class Mail, $0[b] | 580 | 72.0%(2.1%) | 290 | 42.3%(3.7%) |
| First Class Mail, $5[b] | 570 | 84.6%(1.8%) | 260 | 57.4%(3.4%) |
| First Class Mail, $10[b] | 520 | 80.3%(2.3%) | 300 | 61.9%(3.2%) |
| First Class Mail, $15[b] | 500 | 84.5%(2.0%) | 310 | 62.8%(4.1%) |
| First Class Mail, $20[b] | 580 | 82.5%(2.2%) | 290 | 62.9%(3.5%) |

Source: National Household Education Survey Field Test, 2011.
Notes: Standard errors are shown in parentheses. For all cases the first topical follow-up mailing was sent through First Class Mail with no incentive. [a] The second topical follow-up mailing was sent through First Class Mail with either $5 or $15 incentive. [b] The second topical follow-up mailing was sent through Priority Mail with no incentive.
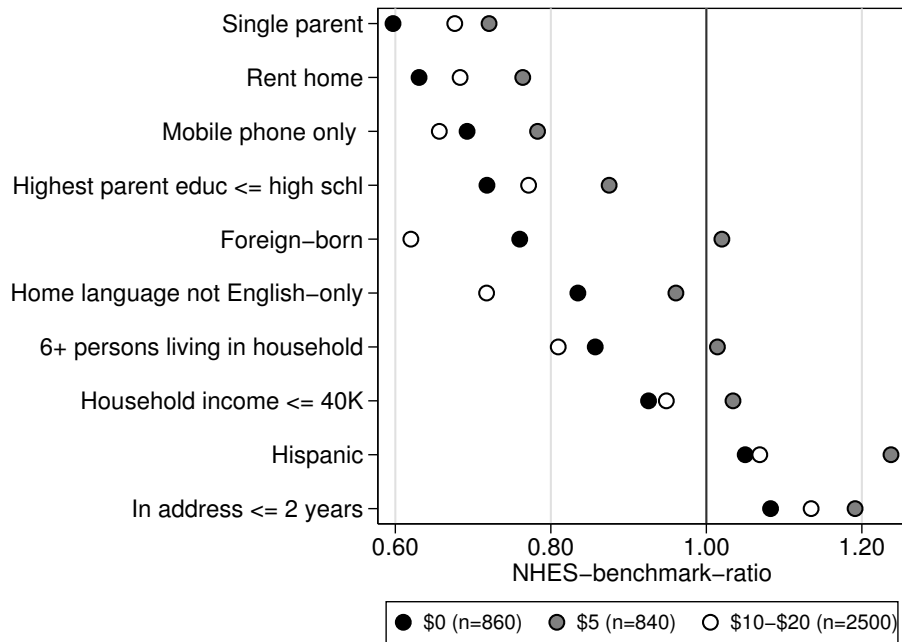


*Figure 5.* Ratios of NHES estimates to benchmarks for demographic and SES characteristics, by topical incentive treatments.

Sources: National Household Education Survey Field Test, 2011; American Community Survey, 2010; National Health Interview Survey, 2011. Notes: The Screener-Early cases that were sent the initial topical mailing via Priority Mail are excluded from the analysis, and the remaining Screener-Early cases are re-weighted to the distribution of Screener-Early versus Screener-Late cases among all the screener respondents. For all the cases in this analysis, the first and second topical follow-up mailings were sent with no incentive, via First Class Mail and Priority Mail respectively.

between screener incentives and topical incentives on sample composition.

Next, we consider the effects of monetary incentives and mailing conditions on the cost of the data collection, specifically the direct cost of mailing and of the monetary incentives themselves. Because other treatments with cost implications (e.g., screener version) varied across cases, we examined the data collection costs for various incentive groups by calculating the unit cost – the average data collection cost across both phases per topical complete. Table 3 shows that the $5

screener-phase treatment drove up the unit cost by approximately 20 percent compared to the $2 treatment. This is true for both the screener-phase cost and the overall cost across both phases. This is partly due to the large amount of screening required to identify the target subpopulation – households with children. For the topical phase, we used the five-group comparison described above. The topical-phase unit cost for the $5 incentive treatment is approximately the same as that for the no incentive group because offering $5 increased the topical response rate substantially.

*Table 3* Effects of incentive treatments on data collection costs

|  | Ratio of unit costs |
| --- | --- |
| Screener treatment | |
|   Screener-phase cost: $5 versus $2 | 1.2 |
|   Overall costs across both phases: $5 versus $2 | 1.2 |
| Topical treatment | |
|   Topical-phase costs: $5 versus $0 | 1.0 |
|   Topical-phase costs: $10 versus $0 | 1.2 |
|   Topical-phase costs: $15 versus $0 | 1.4 |
|   Topical-phase costs: $20 versus $0 | 1.6 |

Source: National Household Education Survey Field Test, 2011.

## 5 Discussion

The "NHES: 2011 Field Test" is the first investigation of the effects of prepaid monetary incentives in a two-phase postal study targeting a specific subpopulation. A mail screener is used to identify the target subpopulation, followed by a mail topical questionnaire that collects the measures of interest. In this design, the second-phase data collection quickly follows the first-phase administration, as opposed to a longitudinal survey where the lag between interviews is generally longer. The research reveals that prepaid monetary incentives are effective in increasing response and affect sample composition.

While the general finding of the usefulness of prepaid monetary incentives holds, the effects of incentives on the overall yield, sample composition and data collection costs do not directly follow the patterns in single-phase studies. In particular, lower incentive levels at both phases have distinct advantages in terms of yield, sample composition, and cost. We observed that the higher incentive levels ($10 or more) at the second phase did not increase the topical response rate beyond that attained by the modest level of $5. Based on the single-phase survey literature, we would have predicted that the higher incentive levels might increase the response rate further (with a diminishing rate of return). In addition, the sample composition associated with the $5 treatment aligns more closely with the benchmarks and represents minority and/or lower SES groups better. Furthermore, the data collection cost per topical complete is the same for the topical $5 group and topical no incentive group. This finding raises an important point that prepaid monetary incentives, if used appropriately, can improve response rate and sample composition without necessarily driving up the data collection cost.

There is no evidence of any interaction effects between the screener and topical incentives. The $5 screener incentive did not condition the screener respondents to expect incentives at the topical phase – when no incentive was offered at the topical phase, the topical response rate for the screener $5 group was not lower than that for the screener $2 group. On the other hand, the positive effect of the screener $5 treatment at the screener phase (compared to the $2 treatment) did not carry over to the topical phase. One explanation is that the additional response the $5 screener treatment induced is from respondents who were less willing to participate and thus the

overall effect was dissipated over the two phases. From a theoretical perspective, social exchange theory as applied in single-phase surveys seems to be a good explanation of the effectiveness of prepaid monetary incentives in two-phase mail surveys. Token prepaid monetary incentives generate higher response rates overall and within key subgroups that tend to respond at lower rates in surveys. We also have evidence that suggests that higher incentive levels such as those used in the topical experiments can produce less desirable results in terms of sample composition and no greater overall yield. This may be because the higher incentive levels frame the request to respond so that it is interpreted as an economic rather than social exchange.

Our research has some limitations. First, the variation in the child weights used for calculating topical response rates limited the statistical power of some analyses (e.g., the comparison between the $5 and $15 treatments for the final topical mailing). Second, the findings about the topical Priority Mail treatment and the incentive treatments for final topical mailing were based on the Screener-Early cases and do not necessarily generalize to all screener respondents. Third, all the cases were offered either $2 or $5 at the screener phase because the single-phase literature had demonstrated the effect of small token incentives. As a result, we cannot measure the impact of the $2 (versus no incentive) on the overall response rate across both phases. Similarly, if no incentive was included as a treatment group for the final topical mailing, we would be able to gauge the impact of the $5 compared to no incentive. Fourth, only a limited number of variables were included in the evaluation of sample composition. Finally, we used the ACS and NHIS data to obtain the benchmarks for sample composition. Although the ACS and NHIS are large-scale surveys sponsored by US government agencies, they are not censuses and thus these benchmark estimates themselves have sampling error.

It is also important to recognize that features of the "NHES: 2011 Field Test" may limit its applicability to other surveys. For example, the particular subpopulation was households with children. Having the US Department of Education as the sponsor and using envelopes featuring the US Department of Education sponsorship seal probably increased the topic salience to the target subpopulation. It is possible that the effect in surveys with commercial or university sponsors might be different. While we believe that

many of the results shown in this paper should generalize, further studies are needed to determine the extent to which these findings apply to other target populations and to surveys with different sponsors and topics.

# References

The American Association for Public Opinion Research (AAPOR). (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 7th edition. The American Association for Public Opinion Research (AAPOR).

AAPOR Cell Phone Task Force. (2010). New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers. http://www.aapor.org/Cell_Phone_Task_Force_Report.htm.

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., & Yamamoto, K. (1992). An Experiment in Monetary Incentives. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 393-398.

Biner, P. M. & Heath J. K. (1994). The Interactive Effects of Monetary Incentive Justification and Questionnaire Length on Mail Survey Response Rates. *Psychology and Marketing*, 11, 483-492.

Brick, J. M., Williams, D., & Montaquila, J. (2011). Address-based Sampling for Subpopulation Surveys. *Public Opinion Quarterly*, 75, 409-428.

Brick, J. M., Montaquila, J., Hagedorn, M. C., Roth, S. B., & Chapman, C. (2005). Implications for RDD Design from an Incentive Experiment. *Journal of Official Statistics*, 21, 571-589.

Cantor, D., O'Hare, B., & O'Connor, K. (2008). The Use of Monetary Incentives to Reduce Non-Response in Random Digit Dial Telephone Surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology*. New York: Wiley.

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly*, 57, 62-79.

Creighton, K. P., King, K. E., & Martin, E. A. (2007). The Use of Monetary Incentives in Census Bureau Longitudinal Surveys. *Survey Methodology* research report series, no. 2007-2. Washington, DC: U.S. Census Bureau.

Dillman, D. A. (1978). *Mail and Telephone Surveys: The Total Design Method.* New York: John Wiley and Sons.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, Third Edition. Hoboken, NJ: John Wiley and Sons.

Gelman, A., Stevens, M., & Chen, V. (2002). *Regression Modeling and Meta-Analysis for Decision Making: A Cost-Benefit Analysis of Incentives in Telephone Surveys.* Journal of Business and Economic Statistics, 21, 213-225.

Gouldner, A. W. (1960). *The Norm of Reciprocity: A Preliminary Statement.* American Journal of Sociology, 25, 161-178.

Griffin, D.H., Fisher, D.P., & Morgan, M.T. (2001). *Testing an Internet Response Option for the American Community Survey.* Paper presented at the 56th Annual Conference of American Association for Public Opinion Research, Montreal, Quebec, Canada.

Griffin, J., Simon, A. B., Hulbert, E., Stevenson, J., Grill, J., Noorbaloochi, S., & Partin, M. (2011). *A Comparison of Small Monetary Incentives to Convert Survey Non-Respondents: A Randomized Control Trial.* BMC Medical Research Methodology, 11:81.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). *Understanding The Decision To Participate In A Survey.* Public Opinion Quarterly, 56, 475-495.

Groves, R. M., & Couper, M. (1998). *Nonresponse in Household Interview Surveys.* New York: John Wiley and Sons.

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

Groves, R.M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta, G.P., Nelson, L. (2006). *Experiments in Producing Nonresponse Bias.* Public Opinion Quarterly, 70, 720-736.

Lesser, V.M., Dillman, D.A., Carlson, J., Lorenz, F., Mason, R., & Willits, F. (2001) *Quantifying the Influence of Incentives on Mail Survey Response Rates and Nonresponse Bias.* Paper presented at the annual meeting of the American Statistical Association, Atlanta, GA.

Han, D., Cantor, D., & Brick, P. (2010). *Findings from a Two-Phase Mail Survey for a Study of Veterans.* Paper presented at the 65th Annual Conference of the American Association for Public Opinion Research, Chicago, IL.

Iannacchione, V.G. (2011). *The Changing Role of Address-Based Sampling in Survey Research.* Public Opinion Quarterly, 75, 556-575.

James, J. M. & Bolstein, R. (1992). *Large Monetary Incentives and Their Effect on Mail Survey Response Rates.* Public Opinion Quarterly, 56, 442-453.

Juster, F. T. & Suzman, R. (1995). *An Overview of the Health and Retirement Study.* Journal of Human Re-sources, 30, S7-S56.

Link, M., Battaglia, M.P., Frankel, M.R., Osborn, L., & Mokdad, A. H. (2008). *A Comparison of Address-Based Sampling (ABS) versus Random Digit Dialing (RDD) for General Population Surveys.* Public Opinion Quar-terly, 72, 6-27.

Martin, E., Abreu, D., & Winters, F. (2001). *Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation.* Journal of Official Statistics, 17, 267-284.

Montaquila, J.M., Brick, J.M., & Kim, K. (2012). *Transitioning from RDD to ABS in the National Household Education Surveys Program.* Paper presented at the Joint Statistical Meetings, San Diego, CA.

Petrolia, D. R. & Bhattacharjee, S. (2009). *Revisiting Incentive Effects: Evidence from a Random Sample Mail Survey on Consumer Preferences for Fuel Ethanol.* Public Opinion Quarterly, 73, 537-550.

Picavet, H. (2001). *National Health Surveys by Mail or Home Interview: Effects on Response.* Journal of Epidemiology and Community Health, 55, 408-413.

Singer, E., Van Hoewyk, J., Maher, M. P. (1998). *Does the Payment of Incentives Create Expectation Effects*? Public Opinion Quarterly, 62, 152-164.

Singer, E. (2002). *The Use of Incentives to Reduce Nonresponse in Household Surveys.* In R.M. Groves, D.A.

Dillman, J.L. Eltinge, & R.J.A. Little (Eds.), *Nonresponse in Household Interview Surveys*, New York: John Wiley and Sons.

Singer, E. & Kulka, R. A. (2002). Paying Respondents for Survey Participation. In M. Ver Ploeg, R. A. Moffitt, & C. F. Citro (Eds.), Studies of Welfare Populations: Data Collection and Research Issues, Washington, D.C.: National Academy Press.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving Survey Response: Lessons learned from the European Social Survey.* New York: John Wiley and Sons.

Trussell, N. & Lavrakas, P.J. (2004). *The Influence of Incremental Increases in Token Cash Incentives on Mail Survey Response: Is There an Optimal Amount?* Public Opinion Quarterly, 68, 349-367.

Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). *Charities, No, Lotteries, No, Cash, Yes: Main Effects and Interactions in a Canadian Incentives Experiment.* Paper presented at the Survey Non-Response Session of the Fourth International Social Science Methodology Conference, University of Essex, Institute for the Social Sciences, Colchester, UK.

Williams, D., Brick, J. M., Montaquila J. M., & Han, D. (2012). *Alternative Questionnaire Effects on Response in Mail Surveys.* Paper presented at the 67th Annual Meeting of the American Association for Public Opinion Research, Orlando, FL.

Yu, J. & Cooper, H. (1983). *A Quantitative Review of Research Design Effects on Response Rates to Questionnaires.* Journal of Marketing Research, 20, 36-44.