Return to Sender: Improving Response Rates in Two-stage Mail Surveys

Saida Mamedova
American Institutes for Research
smamedova@air.org

Cameron McPhee
American Institutes for Research
cmcphee@air.org

1. Introduction

Since 1991, the National Center for Education Statistics (NCES) has used the National Household Education Surveys Program (NHES) to collect data on such topics as early childhood care and education, children's readiness for school, and parent involvement in education. Surveys were conducted approximately every other year from 1991 through 2007, and each of these prior administrations used random digit dial (RDD) sampling and telephone data collection from landline telephones only. Telephone interviews were conducted using computer-assisted telephone interviewing (CATI) to accommodate the survey's complex skip patterns and automated within-household sampling techniques. However, like most RDD surveys, NHES response rates have been declining over time and the increase in households converting from landlines to cell phone-only service has raised concerns about population coverage.

In an effort to address these concerns, NCES opted to redesign the methodology of the NHES data collection and after a smaller-scale feasibility pilot test in 2009 in January 2011, began a large-scale methodological field test of a redesigned multi-mode survey on a national sample of approximately 60,000 addresses in the United States. The primary data collection mode used in the redesigned NHES was a two-phase self-administered mail survey. Included in the field test were multiple embedded experiments intended to determine how to maximize response rates and population coverage, including the use of pre-notice letters, differential incentive levels, different versions of the first-stage screener instrument, and different mailing methods.

This paper explores the effects of the 2011 Filed Test experiments on the response rates. The analyses described below illustrate that sending a pre-notice letter, mailing the second non-response follow-up via Federal Express, and increasing monetary incentives had a statistically significant positive impact on the response rate in the first-stage of data collection. Second-stage differences in response rates were related to both first and second-stage experimental treatments including incentive levels, magnet receipt, and questionnaire customization. Interaction effects between the first-stage and second-stage incentives were also observed, which may suggest that the respondents make the connection between the two stages of the survey, rather than viewing them as separate inquiries.

2. Methodological Design and Justification

In the 2011 NHES Field Test, two large samples were drawn: The primary sample of approximately 41,000 was used to test the effects of the methodological experiments on a nationally-representative sample of households (identified throughout this paper as the "National" sample). Another smaller, non-random, sample was used to test the Spanish-language materials on a targeted Hispanic population. This paper will focus on the analysis of only the National population and therefore further discussion of the experiments will include only those experiments carried out within the National sample.

The first-stage of the survey (the Screener) was designed to screen households for the presence of a child or youth twenty years old or younger (a specific description of the eligibility requirements are discussed in detail below). If the household had at least one eligible child, one child was sampled and the household became eligible for the

second-stage (Topical) questionnaire that asked the respondent detailed questions about childrearing, youth development, and education.

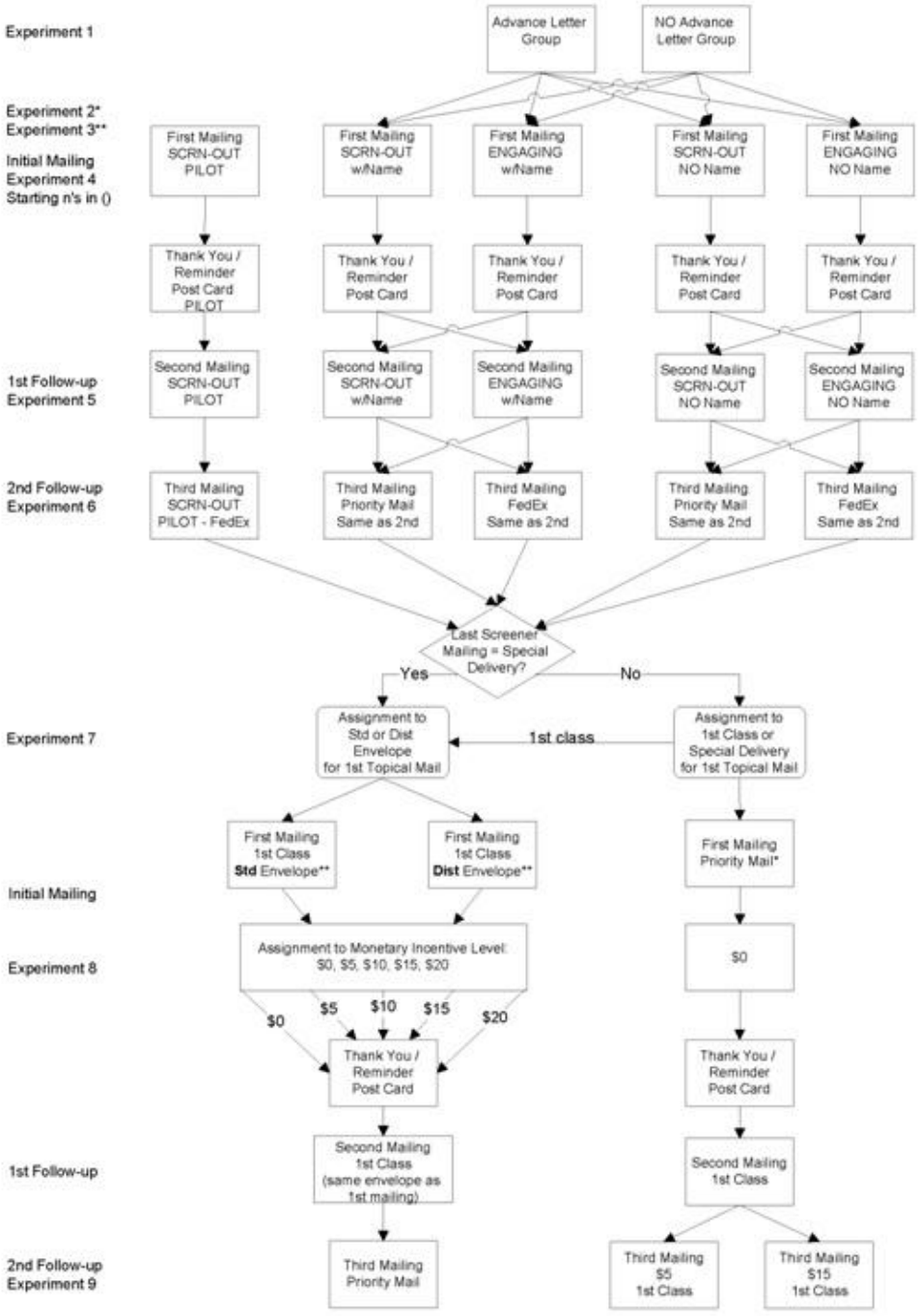
Figure 1 illustrates the experimental design of the 2011 NHES Field Test. The experimental design involved the random pre-designation of addresses to particular treatment groups. An overview of the mailing procedures is outlined below with more detail about the experiments following. Sampled households were mailed an initial screener package which contained a cover letter, screener questionnaire, postage-paid return envelope, varying levels of incentives, and, for a subsample of households, a refrigerator magnet with the Department of Education (DoE) logo. A subsample of 1,500 households was randomly assigned to receive an advance letter 1-2 weeks prior to the first screener mailing in order to test the effect of an advance letter on response rates. The households received a letter describing the study and alerting them to the coming questionnaire. All screener survey instruments contained a question asking whether any children age 20 or younger lived in the household. If there were no children age 20 or younger living in the household, respondents were asked to check a box indicating this and return the form. If there were children or youth age 20 or younger living in the household, respondents were asked to enumerate each child and provide the child's age, sex, school enrollment status, and, if the child was enrolled in school, their grade level and return the form. Approximately two weeks after the initial screener mailing package was sent, households were sent a "Thank you"/Reminder postcard. Non-responding households were then sent a first follow-up mailing package which contained a cover letter, replacement screener questionnaire, and postage-paid return envelope. Households that did not respond to the initial screener mailing or first follow-up mailing were sent a second follow-up mailing that contained a cover letter, replacement screener questionnaire, and postage-paid return envelope. The second follow-up mailing was mailed using one of two priority mailing services[1].

Similar to the first (screener) stage of the NHES: 2011 Field Test, the second (topical) stage of the Field Test included several embedded experiments such as testing the effect of different mailing envelopes, mailing methods, and incentive levels on topical response rates as well as testing different questionnaire wording through the use of a split panel test of two different questionnaires for each topical survey. In households whose returned screener questionnaires indicated the presence of at least one eligible child, one child was selected for a topical follow-up survey. Eligible children fell into two categories: (1) children ages 0 to 6 and not yet enrolled in kindergarten were eligible to receive the Early Childhood Program Participation (ECPP) topical survey and (2) children age 20 or under and enrolled in public or private school for kindergarten through twelfth grade[2] were eligible for the Parent and Family Involvement in Education (PFI) topical survey. In households with multiple children, one child was randomly selected as the focal child for a topical survey in order to minimize burden on respondents[3].

The initial topical mailing package contained a cover letter, topical questionnaire, and postage-paid return envelope. The initial topical mailing was sent via one of two USPS mailing methods (first-class or priority mail). Questionnaires sent by first-class mail were sent in one of two types of envelopes (described in more detail below). Households were also randomly assigned to receive a prepaid monetary incentive of varying amounts with the initial topical mailing[4]. Approximately two weeks after the initial topical mailing, a "Thank you"/Reminder postcard was sent to all households. Non-responding households were then sent a topical first follow-up mailing which contained a cover letter, replacement questionnaire, and postage-paid return envelope. Households that did not respond to the first topical follow-up mailing were sent a second topical follow-up mailing that contained a cover letter, replacement questionnaire, and postage-paid return envelope. For a subset of households that did not receive a monetary incentive with the initial mailing, the second follow-up mailing also contained an incentive.

Figure 1. Experimental design of the NHES 2011 Field Test.

[1] The screener stage had a third non-response follow-up via telephone (CATI) interviewing. Although having higher percentages of non-Whites within the CATI responders, CATI follow-up was able to attract only about a two percent increase in response rate in the National sample overall. This paper examines only the mail mode and, therefore, for the purpose of this paper, CATI responders are considered survey non-responders at the screener, and ineligible for the topical, second-stage mailing.
[2] Homeschooled children were not eligible to receive a PFI survey in the NHES: 2011 Field Test.
[3] Children eligible for the ECPP were sampled at a higher rate within households than school-aged children eligible for the PFI in order to ensure a representative sample of younger children.
[4] Households receiving their initial topical questionnaire via USPS Priority Mail were not eligible to receive a monetary incentive with the initial mailing.



*Experiment 2: Addresses will be randomly assigned to monetary incentive levels for the initial screener mailing

**Experiment 3: Addresses will be randomly assigned to receive a token magnet

***Experiment 10: For each topical survey two versions of each will be fielded for NHES:2011 Field Test. Each version will be randomly distributed among the experimental conditions

Note: Numbers included reflect 10% address ineligibility rate.

Screener

The Screener stage experiments included variations of the incentive levels, envelope contents, mailing methods, and differences in the questionnaire design. The experiments and the theoretical background for inclusion of these experiments are discussed in the next paragraphs.

Pre-notice letter: Approximately 1,500 addresses were randomly sampled to receive an advance letter prior to the first screener mailing. Although the majority of literature on advance letters examines the letters in the context of telephone surveys, the results tend to indicate that pre-notice letters improve response rates (Mann 2005, Link and Mokdad 2005). Some research suggests that in mail surveys, although not affecting the overall willingness to respond, pre-notice letters accelerate the response; with questionnaires returned earlier (Wright 1995). The experiment included in the NHES Field Test was designed to test the application of these findings to a national sample of households in a two-stage, federally sponsored mail survey.

Incentives: The initial mailing of the Screener questionnaires included a $2.00 or $5.00 prepaid incentive. Approximately 18,100 cases of the National sample were randomly assigned $5.00; the remainder received $2.00 (the amount used in prior administrations of the NHES as well as in the 2009 pilot test). The use of incentives has been consistently associated with higher response rates in mail surveys (Petrolia and Bhattacharjee 2009; Lesser, Dillman, Carlson, Lorenz, Mason, and Willits 2001; Church 1993; Armstrong 1975). Lesser et al. (2001) examined the results from eight experiments that included financial incentives ranging from no incentive to $5. Their analysis found that the receipt of financial incentives was associated with higher response, but the higher levels of incentives did not necessarily yield higher response. The NHES Field Test experiment was designed to examine whether prepaid incentives behave similarly in a two-stage survey; and if so, does the effect of the incentive at the screener stage carry over to the second stage or do incentives at the first stage interact with second-stage incentives to influence the propensity of response.

Magnet: Approximately 1,900 randomly assigned households received a magnet with the Department of Education (DOE) logo on it in the initial mailing of the Screener questionnaire. The inclusion of the magnet was designed to address concerns that households may discard the mailings without opening the envelopes. It was hypothesized that including a heavier item of an unusual shape may prompt households to open the questionnaires at a higher rate and, therefore, improve response rates. There is minimal literature about the use of ―items" in this capacity. However, Willimack et al. (1995) concluded that nonmonetary incentives increased response rates, mostly through a reduction in refusal rates, for face-to-face interviews. Ryu, Couper, and Marans (2005) studied the non-cash vs. cash incentives on response rates in the mail survey. Although differing in the type of gift from the 2011 Field Test, Ryu et al. found that a cash incentive yielded higher response than a gift in-kind. NCES wanted to measure whether the Department of Education magnet, coupled with a cash incentive, would improve response rates.

Screener questionnaire versions: Two versions of Screener questionnaires were tested in the National sample. In the initial mailing, approximately 23,700 households received a single page screener that asked only for an enumeration of the children in the household (the ―Screenout" screener); while approximately 17,500 households received a longer screener questionnaire that asked additional education-related questions designed to engage the respondents (the ―Engaging" screener). While research indicates that in interviewer-administered surveys, substantive questions are viewed as a means of establishing rapport and can increase item response rates on sensitive questions (Bradburn, Sudman, and Wansink 2004), Brick (2011) reported that the shorter questionnaires used in the NHES 2009 Pilot study increased the likelihood of response. However, Brick (2011) also observed that the longer version had higher, albeit not statistically significantly higher, response rates from the target population of households with children. Brick supposed that this could be explained by educational items on the Engaging questionnaire being more relevant and attractive for households with children. For the field test (using a larger sample size than the 2009 Pilot study), the experiment was designed to test whether the Engaging questionnaire was more likely to elicit response and if the propensity to respond to the longer questionnaire differed by whether the household had children.

Name of the child: In the National sample, approximately 23,800 households received a screener questionnaire that asked for the names of the child(ren) in the household in addition to their ages and enrollment status[5]. The names were gathered in order to implement the ‗dependent interviewing' (DI) in the second-stage of the survey, which entails using the information gathered in the previous stages to personalize the survey in the subsequent stages. In

[5] Each screener type (―screenout" and ―engaging") had versions that asked for the child(ren)'s names and versions that did not.

the NHES Field Test, households for which we had the child(ren)'s name(s) received a topical questionnaire identifying the sampled child by name in the instructions as well as in many of the questionnaire items. Households that were not asked to provide their child(ren)'s name(s) received questionnaires that identified the focal child by their age, gender, and grade level. Concerns about confidentiality have been known to depress response rates for government-affiliated surveys as Couper et al. (1998) reported in his analysis of the 1990 Decennial Census. Meanwhile, other research (Edwards et al. (2002)) showed that personalization of questionnaires produces higher response rates than controls. The NHES Field Test, therefore, was designed to examine the impact of asking for children's names on response rates at both the screener and topical stages.

Screener questionnaire language: A random subsample (approximately 1,300 cases) of the National sample, received Spanish versions of the Screenout screener questionnaire along with the English version. Half of this subsample received the Spanish questionnaire at every screener mailing; while the other half received Spanish forms only with the 2nd and 3rd mailings (the two nonresponse follow-up mailings). Wobus and de la Puente (1995) used data from the 1993 Spanish Forms Availability Test, a telephone debriefing exercise, to analyze the reaction of the non-Hispanic responders towards receiving the Spanish-language 1990 census form. They report that, while 62 percent of non-Hispanics thought that it was either a "good idea" or "thought nothing of it", 12 percent of non-Hispanics thought it was a bad idea. This experiment was indented to test the impact of receiving Spanish-language materials on response rates for households in areas that are not necessarily linguistically diverse.

Form Switching: In the National sample, those households receiving exclusively English questionnaires were randomly assigned to either receive the same screener version ("screenout" or "engaging") at every screener mailing wave or they were assigned to receive the opposite version from the initial mailing in the 2nd and 3rd mailings. Although, Rendtel and Harms (2009) suggest that people respond better to consistent questionnaires, Brick (2011) suggests, based on the results from the NHES 2009 Pilot study, that households with children are more likely to respond to the "*engaging*" questionnaire at non-response follow-up. The form-switching experiment included in the 2011 Field Test was designed to test whether non-responders to the shorter "screenout" questionnaire would be more likely to respond to the follow-up if they were sent a longer, "engaging" questionnaire, and similarly, if the non-responders to the engaging survey sent at the first-mailing are more likely to respond when sent the shorter, "screenout", questionnaire.

Mailing method (at 2nd non-response follow-up, 3rd wave): Two rush delivery methods were randomly pre-assigned to be tested at the final, wave 3, screener mailing. Those were USPS Priority mail and Federal Express (FedEx). Studying a low-income population in the state of Washington, Gibson et al. (1999) reported that certified mail in the third mailing brought in over twenty five percent of non-respondents and was more effective than the priority mail, regardless of the cash incentive tested. We wanted to test whether this finding held up in the general national population in a federal education survey[6].

2009 Pilot Control: In order to have a control against which to measure any improvement in response rates over the 2009 Pilot study, 5,000 households were randomly chosen to receive the same set of treatments that were used in 2009. This group received a "screenout" questionnaire at all three mailing waves (this questionnaire asked for the Child's name but did not include a question asking the household to provide their phone number), no Pre-notice letter, a $2 prepaid incentive, no DoE magnet, and FedEx mailing at the 3rd wave. Given the non-experimental nature of the individual treatment assignments within the Pilot Control group, these cases are excluded from the analyses presented in this paper.

Topical

The Topical stage experiments focused on testing the questionnaire design, incentive amounts and mailing method for the second-stage of the survey. The experiments and the theoretical background for inclusion of these experiments are discussed in the next paragraphs.

Mailing method: Two mailing methods, USPS Priority mail and USPS First class mail, were assigned for the initial Topical questionnaire mailing. After reviewing ninety eight methodological studies on mail surveys, Heberlein and

[6] Approximately 4,600 addresses were identified as unavailable for FedEx delivery. Regardless of the treatment to which these addresses were initially assigned, all such addresses were mailed the 3rd screener via Priority Mail. These addresses are identified independently in the analyses below so as not to confound the effect of mailing type with region or locale.

Baumgartner (1978) concluded that the use of special mailing procedure, such as certified mail or special delivery, increases the response rates even after controlling for other factors. It was therefore hypothesized that sending the first topical questionnaire via Priority Mail would elicit higher response rates in the topical stage.

Mailing envelop: Those households that received the First class mailing were also assigned to one of two types of envelops – Distinctive and Standard. The distinctive envelope was yellow, rather than white, in order to look different than the first-stage (screener) envelops. The Standard envelope was the same white envelop in which the screener questionnaires were mailed. The hypothesis was that households may be more likely to distinguish the distinctive envelope as a new 2nd questionnaire rather than a duplicate of the screener questionnaire and would, therefore, be more likely to open the envelope. The Distinctive, yellow, envelop was also meant to attract more attention from the recipients in general.

Incentives: Two incentive experiments were included as part of the Topical questionnaire mailing. The first tested the effectiveness of various prepaid incentives included with the first topical mailing. Households were assigned to receive $0, $5, $10, $15, or $20 with the first topical mailing. In their analysis of data from the 2009 NHES Pilot test Brick et al. (2011), found that, after controlling for incentives used in the first stage, the $15 dollar prepaid cash incentive sent in the initial mailing at the second stage of the survey was associated with higher response rates compared to the control group (no incentive) and the group that received a prepaid $5 cash incentive. The $5 incentive was associated with a nominal increase in response relative to the control group but this difference was not statistically significant. In addition to the experiment described above, some households that did not receive an incentive with the first mailing were assigned to receive either $5 or $15 with the final non-response follow-up (the 3rd mailing). This experiment was designed to evaluate the effectiveness of sending an incentive only to harder to reach households as opposed to all households in the initial mailing. Assignment to this treatment, however, was not completely random. This experiment will not be examined in this paper, however McPhee and Hastedt (2012) concluded that even fifteen dollars sent in the final mailing does not motivate the late topical responders to respond at the same rate as if given the money in the beginning.

Questionnaire wording: Questionnaire wording was tested via a split panel design. The primary, "main", form included questions worded similarly to the wording in the 2009 Pilot test. The "alternate" form had questions that had been revised based on the results of cognitive interviewing conducted in 2010. In the National sample, about 5,400 cases received the "main" form and about 2,300 cases received the "alternate" version. The analyses described below do not examine the specific wording differences, but do control for the two versions in analyses of response rates.

There were two conditions that were assigned in the topical stage of the survey according to the screener response: the type of topical questionnaire (Parent and Family Involvement in Education Survey (PFI) and Early Childhood Program Participation Survey (ECPP)) and the language of the questionnaire (English and Spanish). As noted above, the households in which the sampled child was zero to six years of age and was not in kindergarten were sent the ECPP survey. The households in which the sampled child was three to twenty years of age and was enrolled in kindergarten through 12th grade were sent the PFI questionnaire[7]. If the household received and mailed back the Spanish version of the screener or it was requested by telephone, a Spanish topical survey was sent to the household. Topical questionnaire type and language are not analyzed in this paper as experimental treatments, but are included in bi-variate analyses.

3. Methodology

All estimates in this paper are based on the National sample of households excluding addresses that were identified as ineligible after mailing. Addresses were identified as ineligible for the screener mailing if the questionnaire was returned to sender as either undeliverable or a non-residential address. Cases were eligible for the topical mailing if they met the following criteria:

- The address was eligible for the screener

[7] Following is the representation of the assignment to the second-stage questionnaires:

	Child's Age										
Child's Enrollment	0	1	2	3	4	5	6	7	8	//	20
K-12	ECPP	ECPP	ECPP	ECPP	PFI	PFI	PFI	PFI	PFI	PFI	PFI
Not K-12	ECPP	ECPP	ECPP	ECPP	ECPP	ECPP	ECPP	PFI	PFI	PFI	PFI

- The screener was returned and completed such that a child could be sampled from the information provided
- There was at least one eligible child in the household

As noted above, the analysis excludes 2009 Pilot cases because the group was not randomly assigned to treatment conditions independently. All analyses are weighted by the address-level base weights (HBWT and HBWT1--HBWT80) which sum to the number of residential addresses on the sampling frame[8].

The first section below presents bi-variate statistics of response rates by experimental conditions at both the screener and the topical stages. The response rates are presented by mailing wave and rates are cumulative. Since each non-response follow-up mailing adds cost, it was important to determine if any of the experimental treatments increased response overall and at the earlier mailing waves. To examine the impact of the experimental treatments on response time, households were classified according to the "wave" in which they returned a completed screener questionnaire. Wave 1 responders were those who responded to the initial mailing, wave 2 responders were those who responded to the initial or first follow-up (second) mailing, and wave 3 responders were those who responded to any of the screener mailings. In order to classify households whose completed screener questionnaire was returned on a date close to one of the follow-up mailings, the date the questionnaire was returned (screener or topical) was compared to the date the most recent mailing was sent. If a questionnaire was received within three days of the date of the second mailing (the first non-response follow-up), the household was considered a "wave 1 respondent"; if the questionnaire was received within three days of the date of the third mailing (the second non-response follow-up), the household was considered a "wave 2 respondent"; otherwise if the questionnaire was received more than three days after the third mailing, but within the field period, the household was considered a "wave 3 respondent." This decision was made based on the assumption that a household was unlikely to have received and returned a non-response follow-up form within three days of the form being mailed out. Response wave was calculated for both screener and topical questionnaire respondents.

After discussing the relationships between response rates and experimental treatments within the context of bi-variate statistics, the second part of the analyses uses logistic regression to model response propensity controlling for the different experimental conditions. The initial set of logistic regression models examines the impact of screener experimental treatments on the national population. Next the impact of the screener experiments on topical response propensity is examined. In the final analysis, we specify a model that explores the linkages between the two mailing stages on topical response. We conclude by discussing the implications of this analysis in the context of the two-stage mail survey.

Bi-variate analyses were conducted in SAS-callable SUDAAN using the PROC CROSSTAB procedure and the logistic regression analyses were run using the PROC RLOGIST procedure. The following 4 sets of logistic regression models were included in the analyses. Results are presented below:

1. Cumulative screener response at each mailing wave. Independent variables include all screener experimental treatments and screener treatment interactions.
2. Cumulative topical response at each topical mailing wave. Independent variables include all screener experimental treatments, screener response wave, and screener experimental treatment interactions.
3. Cumulative topical response at each topical mailing wave. Independent variables include all topical experimental treatments, and screener response wave.
4. Cumulative Topical response at each topical mailing wave. Independent variables include all screener experimental treatments, screener response wave, screener experimental treatment interactions, topical experimental treatments, and interactions between the screener and topical stages' incentive levels.

In the next section, we present the results of the bi-variate and multi-variate response rate analyses for each mailing stage.

4. Analysis Results

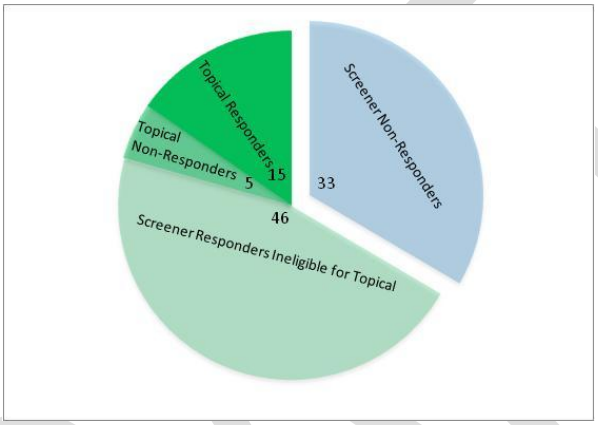
I. Overview

[8] Since households were only sent one topical questionnaire, regardless of the number of children in the household, our analyses of response rates for both the screener and the topical was at the household level and was, therefore, weighted by the household-level base weight.

Overall weighted response rates for the screener and topical stages were 67 and 74 percent of the eligible population respectively, with higher topical response rates explained by the initial responsive contact made at the first stage of the survey (Watson and Wooden 2009). Besides the initial response to the screener survey, the eligibility for the topical survey is determined by households having at least one eligible child from 0 to 20 years of age. The 31 percent of households eligible for the topical questionnaires, although not precisely identically defined, compares very well with the 29 percent of households with children of 0 to 17 years of age reported by the Current Population Survey (Current Population Survey 2011).

Figure 2 shows response and eligibility distributions for the two – topical and screener – stages. The majority of sampled households responded to the screener, as did the majority of households eligible for the topical mailing. Our final target population constitutes a small proportion of our total sample as indicated above.

Figure 2. Response and eligibility distributions for screener and topical stages (percent)[9]



II. Screener response rates: bi-variate analysis

Table A.1 in Appendix A details each of the experiments' response rates discussed below. Receiving a pre-notice letter has a statistically significant differential impact on the response rates in the first mailing. However, these effects disappear, becoming statistically non-significant, by the end of the screener stage. This suggests that receiving a pre-notice letter has a strong immediate effect, precipitating earlier response, but does not impact the final response rate. Asking for the name of the child may slightly depress responses throughout the waves, but this effect was not statistically significant. Excluding those cases located at Fedex undeliverable addresses, FedEx fares positively compared with Priority mail suggesting that FedEx may more effectively attract the attention of non-responding households. The level of incentives shows a significant impact on responsiveness. Households receiving $5 with the initial screener mailing, compared to those receiving $2 were statistically more likely to respond at all three screener mailing waves. Providing a DoE magnet in the initial screener mailing did not improve screener response rates at any of the three screener waves. Finally, the shorter screener survey (Screenout) generally produced better overall final screener response rates than the engaging questionnaire particularly when sent at all 3 mailing waves (the screenout-screenout treatment).

III. Screener response rates: logistic regressions

Table 1 shows the logistic regression models for the screener response rates at each screener mailing wave. In general, the regression results parallel the bi-variate statistics discussed above. In addition, the logistic regression results emphasize that sending the engaging survey at the second and third waves negatively impact the overall cumulative response rates. Compared with the households that received screenout questionnaires throughout all three waves (screenout-screenout), households receiving engaging questionnaires for all three waves (engaging-engaging) or for the second and third mailing waves (screenout-engaging) responded less often. Households that

[9] Address-based weighted estimates of our analytical, national non-Pilot, sample.

received a Spanish-language screener in addition to the English screener at any point in the screener mailing phase responded at the same rate, statistically, as households receiving English screeners at all mailing waves.

Among the screener experiments, the pre-notice letter and five-dollar incentive both show the strongest influence at the initial mailing of the screener questionnaires (odds ratios of 1.25 and 1.31 respectively), diminishing in effect size by the end of the data collection, with the pre-notice letter losing its statistical significance by wave 3. Among all the experiments, the five-dollar incentive has the highest impact on the likelihood of response (odds ratio of 1.23 at the end of the first stage). Mailing by FedEx at the last wave and not changing the shorter questionnaire to the longer one after the first wave (Screenout-Engaging) are similarly beneficial for the overall response rate (odds ratios of 1.14 and 0.88 respectively). There is a slight improvement of the response rate if the longer questionnaire (Engaging-Engaging) is not used at all (odds ratio of 0.91).

Table 1. Logistic Regression: Cumulative Screener Response Rates at each Mailing Wave by Screener Experimental Treatments

	Odds Ratios		
	Model 1 (Dep Var: Screener Completion by Wave 1)	Model 2 (Dep Var: Screener Completion by Wave 2)	Model 3 (Dep Var: Screener Completion by Wave 3)
Screener Experimental Conditions			
Pre-notice Letter			
Pre-notice Letter	1.25 *	1.14 *	1.06
No Pre-notice Letter	--	--	--
Name			
Name	0.96	0.94 *	0.95
No Name	--	--	--
Mailing method			
Control: FedEx Undeliverable	0.65 *	0.71 *	0.70 *
FedEx	1.01	1.01	1.14 *
Priority mail	--	--	--
Incentives			
$5.00 Screener	1.31 *	1.28 *	1.23 *
$2.00 Screener	--	--	--
Magnet			
Magnet	1.02	1.02	1.06
No Magnet	--	--	--
Screener questionnaire version			
Screenout-Engaging	1.00	0.97	0.88 *
Engaging-Engaging	0.97	0.96	0.91 *
Engaging-Sreenout	0.96	0.97	1.01
Spanish Screenout	1.07	1.10	1.07
Screenout-Screenout	--	--	--

-- Reference category.
Significance levels are marked as * when confidence is at least 95 percent (p<0.05)

From previous research there is reason to believe that some experimental treatments may interact with each other, leading one variable to moderate the impact of another on response rates. For example, Knecht et al. (2011) has noted that screener respondents to the 2009 NHES Pilot test were less likely to provide the child(ren)'s name(s) if they received the longer (engaging) screener compared to those who received the shorter (screeneout) questionnaire. It therefore is possible that the impact on overall response rates of *asking* for the child(ren)'s names on response rates may be influenced by the screener type.

The introduction of interaction terms in regression models complicates the interpretation of the odds ratios. Specifically, the odds ratios for the overall categories cannot be interpreted separately from the odds ratios in the interaction terms. For example, in Table 2, the odds ratio for the "child's name" group (0.90) is the odds ratio for asking for the child's name compared to not asking within the reference category of the other independent variable of the interaction term (in this case screener type). In other words, it is the impact on response rates of asking for a child's name compared to not asking for the name for households who received the screenout questionnaire throughout all waves (screenout-screenout). Specifically, these results indicate that for those households receiving the screenout questionnaire at all waves, the odds of responding to the questionnaire are diminished by asking for

the child's name on the questionnaire. Similarly, the screenout-engaging odds ratio in the table (0.81) shows the impact of the change in questionnaire from shorter (screenout) to longer (engaging) compared to receiving the screenout at all waves for those households receiving questionnaires on which the child's name was not solicited (the reference category child's name condition). The odds ratios for the interaction itself compare the odds ratios for the child's name experiment across the different screener questionnaire types. For example in the table below, the Name :: Screenout-Engaging interaction indicates that the negative relationship between asking for the child's name and response rates is weaker for households receiving the screenout questionnaire first and then the engaging questionnaire compared to households that received the screenout questionnaire at all mailing waves.

Table 2. Logistic Regression: Cumulative Screener Response Rates at 3rd Screener Mailing Wave by Screener Experimental Treatments and Interactions

	Odds Ratios	
Screener Experimental Conditions		
Pre-notice Letter		
Pre-notice Letter	1.06	
No Pre-notice Letter	--	
Name		
Name	0.90	*
No Name	--	
Mailing method		
Control: FedEx Undeliverable	0.70	*
FedEx	1.14	*
Priority mail	--	
Incentives		
$5.00 Screener	1.23	*
$2.00 Screener	--	
Magnet		
Magnet	1.06	
No Magnet	--	
Screener questionnaire version		
Screenout-Engaging	0.81	*
Engaging-Engaging	0.90	
Engaging-Sreenout	1.00	
Spanish Screenout	1.10	
Screenout-Screenout	--	
Interactions		
Interaction - Name :: Screenout-Engaging	1.18	*
Interaction - Name :: Engaging-Engaging	1.02	
Interaction - Name :: Engaging-Screenout	1.03	
Interaction - Name :: Spanish Screenout	1.00	
Interaction - Name :: Screenout-Screenout	--	

-- Reference category.
Significance levels are marked as * when confidence is at least 95 percent (p<0.05)

IV. Topical Response Rates: bi-variate analysis

As mentioned earlier, screener respondents were only eligible to receive the topical questionnaires if it was determined from their screener answers that the household had children 20 years of age or younger. Thirty one percent of the screener responders matched this criteria. Since it is possible that screener experimental treatments could impact the likelihood that households with children respond to the second stage, topical, questionnaire, we first examined how the topical response rates varied by screener experimental conditions.

Table A1 in Appendix A details the topical response rate at each topical mailing wave by the screener and topical experimental treatments. The bi-variate analysis suggests that asking for the child(ren)'s name on the screener does produce an increase in topical response rates. This difference appears beginning in wave 2 of the topical mailing suggesting that later responders to topical questionnaires react to seeing the name of the focal child differently than do earlier responders. Excluding responders who lived in the Fedex undeliverable addresses, sending a Fedex in the

last wave of the screener stage improved topical response rates at wave 2, but not at initial or in the final third mailing of the topical[10]. Those households that received two dollars at the screener stage have a statistically significantly higher response rates for all three waves of the topical stage. This will be discussed further with the logistic regression results below. Topical response rates were also positively impacted by the Department of Education magnet sent with the initial screener. It is possible that while the magnet did not improve response rates at the screener stage, households may hold on to the magnet which then serves as a reminder of the study when the topical questionnaire arrives. Receiving a pre-notice letter prior to the screener mailing did not statistically improve response rates at the topical stage, nor did the type of screener questionnaire (screenout or engaging) completed.

Topical response rates also varied depending on the experimental treatments used in the topical mailing stage. There are several reasons to exercise caution when interpreting the bi-variate results of the topical experiments. First, it is likely that some of the topical treatments may be interacting with the first-stage, screener, experiments that are not controlled for here. Additionally, as we will discuss later, the timing of screener response had a significant impact on the likelihood of response at the topical stage and the topical mailing procedures were, in part, determined by the timing of screener response. Despite these complexities, the bi-variate statistics will offer a brief overview of the impact of the topical experimental treatments.

As can be seen in table A.1 in Appendix A the bi-variate analyses shows receiving the topical questionnaire via First class mail in a ―distinctive" envelop impacted the response rates positively in the third wave[11]. Although, the last mail was sent Priority mail for both Distinctive and Standard envelops, the results suggest that getting the Distinctive envelop twice makes a difference regardless of the mailing method for the third mailing. Across the incentive levels, the statistically significant and clearly distinctive effect was for those who did not receive any monetary incentive compared to those who did receive some compensation. Interestingly, recipients of the five dollar incentive in the topical stage had similar final response rates as fifteen and twenty dollar recipients at the close of data collection. The item wording differences between forms had no statistical relationship to response rates at any mailing wave.

V. Topical Response Rates: Logistic Regression

As noted in the literature, second stage response propensity is strongly related to the likelihood of response in stage one. As Watson et al. (2009) points out in their review of the response rates in the panel studies, cooperative respondents are also much more likely to be responsive in the next wave. Conversely, respondents requiring the most effort at the previous wave are much more likely to not respond at all in the current wave (Nicoletti et. al, 2005). The latter research relates to the question of whether the wave three screener responders are likely to be responders at all for the topical stage. In order to answer this question as well as control for the timing of screener response when examining the impact of topical experimental treatments on response propensity, the logistic regression analysis discussed below included the screener wave of response in the logistic regressions as controls.

Table 3 below displays the results of the logistic regressions, examining the impact of the first-stage treatments on the second-stage response rates. The cases that received the screeners asking for the name of the child were sent personalized topical questionnaires in the second stage. The analysis below shows that the personalization does relate to higher response rates after the first wave of topical mail-outs, even after controlling for the screener wave of response (odds ratios of 1.18 at wave 2, and 1.21 at wave 3). Meanwhile, receiving a Department of Education magnet encourages the response throughout all the waves. In fact, the magnet treatment has the largest positive impact on the responsiveness of all screener treatments at the end of the topical stage (odds ratio of 1.40). The five-dollar screener incentive appeared to negatively impact the topical response rates (odds ratios of 0.84 at the end of the topical stage). We will discuss this phenomenon in more detail later. Controlling for FedEx undeliverable addresses, the mailing method used with the 3rd mailing wave was not related to second-stage response rates. Our results do show a highly significant and highly influential impact of the timing of screener responses on the topical response rates across all waves (odds ratios of 2.72 for screener wave 1 responders and 1.55 for screener wave 2 responders compared with the screener wave 3 responders at the end of the topical stage).

[10] This finding should be interpreted with caution since the mailing procedure at the topical stage was not random, but was based on when the household responded to the screener.

[11] The effect of sending the first topical mailing via Priority Mail is not discussed because this treatment was not randomly assigned.

When the relationships are evaluated including the interactions of the first-stage experiments, personalizing the topical questionnaire for the households that got a shorter questionnaire throughout (screnout-screenout) has a significantly positive effect on the final response rate. However, the significant interaction term for Name::Screenout-Engaging suggests that the personalization has less of an impact on response rates for households that received the screenout first followed by the engaging questionnaire compared to those receiving the screenout at all waves.

Table 3. Cumulative Topical Response Rates at each Wave by Screener Experimental Conditions and Interactions

	Odds Ratios			
	Model 1 (Dep Var: Topical Completion by Wave 1)	Model 2 (Dep Var: Topical Completion by Wave 2)	Model 3 (Dep Var: Topical Completion by Wave 3)	Model 4 (Dep Var: Topical Completion by Wave 3 w/ Group 4 and Interactions)
Screener Experimental Conditions				
Pre-notice Letter				
Pre-notice Letter	1.00	0.99	0.93	0.93
No Pre-notice Letter	--	--	--	--
Name				
Name	1.05	1.18 *	1.21 *	1.52 *
No Name	--	--	--	--
Mailing method				
Control: FedEx Undeliverable	0.63 *	0.74 *	0.69 *	0.69 *
FedEx	1.07	1.09	1.02	1.02
Priority mail	--	--	--	--
Incentives				
$5.00 Screener	0.83 *	0.84 *	0.84 *	0.84 *
$2.00 Screener	--	--	--	--
Magnet				
Magnet	1.28 *	1.29 *	1.40 *	1.39 *
No Magnet	--	--	--	--
Screener questionnaire version				
Screenout-Engaging	1.09	0.94	0.89	1.01
Engaging-Engaging	1.07	0.99	0.96	1.14
Engaging-Sreenout	1.05	1.03	0.99	1.11
Spanish Screenout	0.91	0.73 *	0.91	0.80
Screenout-Screenout	--	--	--	--
Interactions				
Interaction - Name :: Screenout-Engaging	†	†	†	0.77
Interaction - Name :: Engaging-Engaging	†	†	†	0.69 *
Interaction - Name :: Engaging-Screenout	†	†	†	0.77
Interaction - Name :: Spanish Screenout	†	†	†	1.00
Interaction - Name :: Screenout-Screenout	†	†	†	--
Screener Wave Response				
by Wave 1	2.26 *	2.65 *	2.72 *	2.73 *
by Wave 2	1.42 *	1.53 *	1.55 *	1.56 *
by Wave 3	--	--	--	--

-- Reference category.
Significance levels are marked as * when confidence is at least 95 percent (p<0.05)

Table 4 shows the effects of the topical experiments on topical response rates controlling for the timing of the screener wave response. As seen from the experimental routing for topical questionnaires (Figure 1), some of the topical experimental conditions were based on the timing of the screener wave response. Responders to the last wave of the screener (late screener responders) were not eligible to receive the initial topical mailing via priority mail. In addition, *all* late screener responders (in addition to a random sample of 1st and 2nd wave screener responders) received one of the five options for first wave topical incentives: $0, $5, $10, $15, or $20. Those screener cases responding before the third wave of the screener (early responders) that *were* randomly assigned to Priority mail did not have a chance to receive $5, $10, $15, or $20 incentives at the topical first wave.[12] Given this experimental routing at the topical stage, it is important that the analysis control for the screener timing of response. Controlling for the timing of the household's response to the screener, the topical incentives sent with the first

[12] Cases receiving their first topical mailing via Priority Mail all received $0 at the topical first wave and were eligible to receive either $5 or $15 at the final non-response follow-up (wave 3) mailing. The results of this experiment are not presented here. Please see McPhee and Hastedt (2012).

mailing are strongly related to the topical response rates. Compared with the households who did not receive any incentive at the first mailing, the group receiving $5 had the highest relative increase (odds ratio of 1.81 at wave 3). Additionally, the $15 and $20 dollar groups are more than twice as likely to respond as $0 group. The effect of incentives is the highest at the initial mailing, but sustains throughout all three mailing waves. Meanwhile neither the mailing method, nor the type of envelop, nor the type of questionnaire show a significant impact on the response rates at any point in the topical stage.

Table 4. Cumulative Topical Response Rates at each Wave by Topical Experimental Conditions and Screener Response Wave

	Odds Ratios		
	Model 1 (Dep Var: Topical Completion by Wave 1)	Model 2 (Dep Var: Topical Completion by Wave 2)	Model 3 (Dep Var: Topical Completion by Wave 3)
Screener Wave Response			
by Wave 1	2.95 *	3.51 *	3.55 *
by Wave 2	1.77 *	1.95 *	1.95 *
by Wave 3	--	--	--
Topical Experimental Conditions			
Mailing envelop			
FC - Distinctive	0.96	1.05	1.02
FC - Standard	0.90	0.92	0.87
Priority mail	--	--	--
Incentives			
$5 Incentive	1.87 *	1.78 *	1.81 *
$10 Incentive	1.99 *	1.92 *	1.92 *
$15 Incentive	2.17 *	2.17 *	2.23 *
$20 Incentive	2.45 *	2.38 *	2.42 *
$0 Incentive	--	--	--
Questionnaire wording			
Main	1.06	1.06	1.02
Alternate	--	--	--

-- Reference category.
Significance levels are marked as * when confidence is at least 95 percent (p<0.05)

The next section examines the impact on topical response rates of all the experimental treatments combined, including interactions between the first stage and second stage incentive levels. As, Laurie and Lynn (2009) note, prior research suggests that payments in the first stages raise expectations in the subsequent waves. This phenomenon is also discussed in the analysis by McPhee and Hastedt (2012).

Table 5 displays the odds ratios for the logistic regression models examining the impact of all the 2011 Field Test experimental treatments on the final topical response rates. The majority of the relationships discussed in the previous sections remained significant in the models below. For this reason, only the new interaction terms, between the screener and topical incentive levels, are discussed here. As was seen in the earlier analysis presented in Table 3, households that received $5 with the initial screener mailing were less likely to respond to the topical questionnaire than households that received $2. It was hypothesized that this may, in fact, be the result of an interaction between the level of incentive sent with the screener and the level sent with the topical. Perhaps, households receiving no money at the topical were less likely to respond if they had been given $5 with the screener compared to $2. However, the regression results presented in Table 5 indicate that while the $5 screener incentive is definitely negatively associated with topical response for those households receiving no money with the topical (odds ratio of 0.83 at wave 3), the interaction terms are non-significant. Second stage incentives impact the final response rates as expected – the greater the incentive the more positive is the response. While not statistically significant, the odds ratios for the interaction terms suggest that with increased power, an interaction between the incentive levels across stages may emerge. This implies that for a two-stage survey giving money in the first stage without plans of giving money in the second stage may harm the response rates. If incentives are planned in both stages, the first stage incentives and second stage incentives should not be considered independently.

Table 5. Topical Response Rates at each Wave by Screener and Topical Experimental Conditions, Interactions, and Screener Wave of Response

	Odds Ratios		
	Model 1 (Dep Var: Topical Completion by Wave 1)	Model 2 (Dep Var: Topical Completion by Wave 2)	Model 3 (Dep Var: Topical Completion by Wave 3)
Screener Experimental Conditions			
Pre-notice Letter			
Prenote Letter	0.99	0.97	0.92
No Prenote Letter	--	--	--
Name			
Name	1.22 *	1.39 *	1.53 *
No Name	--	--	--
Mailing method			
Control: FedEx Undeliverable	0.63 *	0.75 *	0.70 *
FedEx	1.08	1.10	1.03
Priority mail	--	--	--
Incentives			
$5.00 Screener	0.81 *	0.81 *	0.83 *
$2.00 Screener	--	--	--
Magnet			
Magnet	1.29 *	1.30 *	1.41 *
No Magnet	--	--	--
Screener questionnaire version			
Screenout-Engaging	1.16	0.99	1.01
Engaging-Engaging	1.22 *	1.13	1.14
Engaging-Sreenout	1.17	1.13	1.10
Spanish Screenout	0.85	0.68 *	0.81
Screenout-Screenout	--	--	--
Interactions			
Interaction - Name :: Screenout-Engaging	0.91	0.92	0.78
Interaction - Name :: Engaging-Engaging	0.79	0.77	0.70 *
Interaction - Name :: Engaging-Screenout	0.81	0.82	0.78
Interaction - Name :: Spanish Screenout	1.00	1.00	1.00
Interaction - Name :: Screenout-Screenout	--	--	--
Screener Wave Response			
by Wave 1	2.99 *	3.61 *	3.65 *
by Wave 2	1.79 *	1.99 *	2.00 *
by Wave 3	--	--	--
Topical Experimental Conditions			
Mailing envelop			
FC - Distinctive	0.95	1.03	1.01
FC - Standard	0.90	0.92	0.86
Priority mail	--	--	--
Incentives			
$5 Incentive	1.80 *	1.68 *	1.73 *
$10 Incentive	1.83 *	1.85 *	1.92 *
$15 Incentive	2.31 *	2.18 *	2.17 *
$20 Incentive	2.37 *	2.25 *	2.37 *
$0 Incentive	--	--	--
Questionnaire wording			
Main	1.07	1.06	1.03
Alternate	--	--	--
Interactions			
Interaction - $5 Sceener Incentive :: $5 Topical Incentive	1.08	1.14	1.11
Interaction - $5 Sceener Incentive :: $10 Topical Incentive	1.17	1.07	1.01
Interaction - $5 Sceener Incentive :: $15 Topical Incentive	0.88	1.02	1.10
Interaction - $5 Sceener Incentive :: $20 Topical Incentive	1.09	1.14	1.06
Interaction - $5 Sceener Incentive :: $0 Topical Incentive	--	--	--

† Not applicable.

-- Reference category.

Significance levels are marked as * when confidence is at least 95 percent (p<0.05)

5. Conclusion

For the general population, sending a pre-notice letter solicits earlier response to the screener, thereby saving time and money on follow-up mailings. Increasing the prepaid incentive level from $2 to $5 significantly increases the response rates at the screener stage, but may negatively impact response rates at the topical stage. Therefore, the optimal level of incentive in the first stage will always depend on the level of response needed at that stage since the overall response rate is the multiplied response rates of both stages. Our analysis does suggest that if incentives are given at the first stage, it is strongly advisable to give at least that same amount of incentives at the topical stage. This finding reiterates the fact that the decision about the amount sent in the second stage cannot be made in isolation; rather the necessary response rates for both stages must be examined and the incentive levels for each stage determined together.

Our analysis also determined that a shorter screener questionnaire generally elicits better screener response rates than a longer one and mailing the non-response follow-up by FedEx increases the likelihood of response for the general population.

For our target population of households with children, personalizing the topical questionnaire with the child‘s name (collected on the screener) increases the likelihood of response at the 2nd and 3rd topical mailing stages. Since asking for the name on the screener was not shown to decrease screener response significantly, asking for the name benefits the overall response rates. In addition, sending a magnet to the households in the first stage also affects the topical (second stage) response rates positively. The strong impact of the timing of the first-wave response on the second-wave response suggests that those households responding quickly differ from those who respond later, the specifics of which should be examined more closely.

Appendix A

Table A.1 Screener and Topical Response Rates at Each Wave by Screener and Topical Experimental Conditions

	Screener			Topical		
	S Wave 1	S Wave 2	S Wave 3	T Wave 1	T Wave 2	T Wave 3
Total	42.8	55.0	66.6	41.6	64.0	74.2
Screener Experimental Conditions						
Pre-notice Letter						
Pre-notice Letter	48.0	58.0	67.8	42.6	64.6	73.7
No Pre-notice Letter	42.6	54.9	66.5	41.6	64.0	74.2
Name						
Name	42.5	54.3	66.2	41.8	65.3	75.6
No Name	43.3	55.7	67.0	41.4	62.6	72.7
Mailing method						
FedEx	43.9	55.9	68.8	43.0	65.2	74.7
Priority mail	43.5	55.5	65.8	41.7	63.9	74.6
Control: FedEx Undeliverable	33.4	47.2	57.6	30.5	55.7	66.4
Incentives						
$2.00 Screener	39.6	51.9	64.3	43.6	65.6	75.4
$5.00 Screener	46.1	58.1	68.9	39.8	62.5	73.0
Magnet						
Magnet	43.4	55.6	67.9	47.5	69.5	79.8
No Magnet	42.8	55.0	66.5	41.3	63.7	73.8
Screener questionnaire version						
Screenout-Screenout	43.3	55.5	67.7	40.7	64.5	75.0
Screenout-Engaging	43.3	54.8	64.8	42.8	63.5	73.1
Engaging-Engaging	42.4	54.5	65.5	42.0	64.0	74.1
Engaging-Sreenout	42.2	54.8	67.9	41.4	64.6	74.3
Spanish Screenout	44.5	57.2	69.8	39.3	59.9	75.0
Topical Experimental Conditions						
Mailing method						
Priority	†	†	†	38.4	62.2	73.5
First class	†	†	†	43.6	65.1	74.6
Mailing envelop						
FC - Distinctive	†	†	†	44.3	66.4	76.1
FC - Standard	†	†	†	42.9	63.9	73.2
Priority mail	†	†	†	38.4	62.2	73.5
Incentives						
$0 Incentive	†	†	†	36.4	59.8	70.9
$5 Incentive	†	†	†	44.9	66.3	75.8
$10 Incentive	†	†	†	45.3	66.6	75.8
$15 Incentive	†	†	†	47.0	68.9	78.1
$20 Incentive	†	†	†	50.8	71.6	80.1
Questionnaire wording						
Main	†	†	†	42.2	64.5	74.4
Alternate	†	†	†	40.2	62.7	73.5
Questionnaire Type						
ECPP	†	†	†	41.9	65.6	74.6
PFI	†	†	†	41.5	63.3	74.0
Language of Questionnaire						
English	†	†	†	41.7	64.0	74.1
Spanish	†	†	†	22.7 !	63.6 !	81.8

† Not applicable.

! Interpret data with caution (estimates are unstable).

Pilot cases excluded from estimates.

Table A.2 Standard Errors: Screener and Topical Response Rates at Each Wave by Screener and Topical Experimental Conditions

	Screener			Topical		
	S Wave 1	S Wave 2	S Wave 3	T Wave 1	T Wave 2	T Wave 3
Total	0.27	0.27	0.24	0.56	0.51	0.56
Screener Experimental Conditions						
Pre-notice Letter						
Pre-notice Letter	1.35	1.36	1.41	2.62	2.43	2.35
No Pre-notice Letter	0.28	0.27	0.24	0.57	0.53	0.56
Name						
Name	0.35	0.35	0.35	0.70	0.80	0.77
No Name	0.42	0.43	0.41	0.83	0.84	0.80
Mailing method						
FedEx	0.39	0.38	0.35	0.73	0.74	0.74
Priority mail	0.42	0.46	0.40	0.89	0.87	0.83
Control: FedEx Undeliverable	0.80	1.01	1.06	2.25	2.30	2.29
Incentives						
$2.00 Screener	0.35	0.36	0.33	0.98	0.83	0.77
$5.00 Screener	0.40	0.42	0.37	0.76	0.85	0.79
Magnet						
Magnet	1.15	1.25	1.19	2.61	2.29	2.23
No Magnet	0.27	0.26	0.24	0.56	0.51	0.57
Screener questionnaire version						
Screenout-Screenout	0.59	0.62	0.56	1.22	1.23	1.16
Screenout-Engaging	0.49	0.54	0.52	1.20	1.26	1.14
Engaging-Engaging	0.57	0.51	0.49	1.26	1.15	1.08
Engaging-Sreenout	0.62	0.63	0.57	1.06	1.05	0.89
Spanish Screenout	1.61	1.28	1.20	3.03	2.97	2.65
Topical Experimental Conditions						
Mailing method						
Priority	†	†	†	0.97	1.01	0.94
First class	†	†	†	0.73	0.68	0.68
Mailing envelop						
FC - Distinctive	†	†	†	1.11	1.10	0.98
FC - Standard	†	†	†	0.92	0.98	0.94
Priority mail	†	†	†	0.97	1.01	0.94
Incentives						
$0 Incentive	†	†	†	0.83	0.85	0.84
$5 Incentive	†	†	†	1.63	1.54	1.43
$10 Incentive	†	†	†	1.60	1.51	1.40
$15 Incentive	†	†	†	2.07	1.74	1.64
$20 Incentive	†	†	†	1.74	1.57	1.36
Questionnaire wording						
Main	†	†	†	0.70	0.61	0.61
Alternate	†	†	†	0.99	1.08	1.03
Questionnaire Type						
ECPP	†	†	†	1.15	1.10	1.02
PFI	†	†	†	0.65	0.65	0.67
Language of Questionnaire						
English	†	†	†	0.56	0.51	0.56
Spanish	†	†	†	9.03	10.49	8.73

† Not applicable.

Pilot cases excluded from estimates.

References

Armstrong, J.S. (1975). Monetary Incentives in Mail Surveys. *Public Opinion Quarterly, 39(1): 111-116*.

Bradburn, N., Sudman, S., and Wansink, B. (2004). Asking Questions: the Definitive Guide to Questionnaire Design—For Market Research, Political Polls, and Social and Health Questionnaires. Jossey-Bass: San Francisco, CA.

Brick, M.J., William, D., and Montaquila, J. (2011). Address-Based Sampling for Subpopulation Surveys. *Public Opinion Quarterly, 75(3): 409-428*.

Church, A. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly, 57(1): 62-79*.

Couper, M.P., Singer, E., and Kulka R.A. (1998). American Politics Research (January 1998), 26(1): 59-80. Stable URL: http://deepblue.lib.umich.edu/bitstream/2027.42/67793/2/10.1177_1532673X9802600104.pdf

Current Population Survey (2011). America‘s Families and Living Arrangements: 2011. U.S. Census Bureau, Current Population Survey, 2011, Annual Social and Economic Supplement. Internet Release Date: November 2011: Tables AVG1 and F2.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., and Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *British Medical Journal 2002: 324-1183*. Stable URL: http://www.bmj.com/content/324/7347/1183?tab=full

Ford, N.M. (1967) The Advance Letter in Mail Surveys. *Journal of Marketing Research, 4(1): 202-204*. American Marketing Association. Stable URL: http://www.jstor.org/stable/3149368

Gibson, P.J, Koepsell, T.D., Diehr, P., and Hale,C. (1999). Increasing Response Rates for Mailed Surveys of Medicaid Clients and Other Low-income Populations. *American Journal of Epidemiology, 149(11): 1057-1062*. The Johns Hopkins University School of Hygiene and Public Health. Stable URL: http://aje.oxfordjournals.org/content/149/11/1057.full.pdf

Heberlein, A.T., and Baumgartner, R. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. *American Sociological Review, 43(4): pp. 447-462*. Stable URL: http://www.jstor.org/stable/2094771

Knecht, A., Grady, S., and McPhee C. (2011). ―Differences in Respondent Propensity to Disclose Child‘s First Name in the NHES 2009 Pilot Study.” Presentation at the American Association for Public Opinion Research (AAPOR) 66th Annual Conference. Phoenix, AZ. May, 2011.

Laurie, H. and Lynn, P. (2009). The Use of Respondent Incentives on Longitudinal Surveys. Methodology of Longitudinal Surveys. (ed. P. Lynn), Copyright 2006 (John Wiley & Sons, Ltd.). Stable URL: http://www.iser.essex.ac.uk/files/iser_working_papers/2008-42.pdf

Lesser, V.M., Dillman, D.A., Carlson, J., Lorenz, F., Mason, R., and 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. Stable URL: http://www.sesrc.wsu.edu/dillman/papers/2001/QuantifyingTheInfluence.pdf

Link, M.W. and Mokdad A. (2005). Advance Letters as a Means of Improving Respondent Cooperation in Random Digit Dial Studies. Public Opinion Quarterly (Winter 2005) 69 (4): 572-587. Stable URL: http://poq.oxfordjournals.org/content/69/4/572.full

Mann, C.B. (2005). Do Advance Letters Improve Preelection Forecast Accuracy. *Public Opinion Quarterly, 69(4): 561-571*. Stable URL: http://poq.oxfordjournals.org/content/69/4/561.full.pdf

McPhee, C. and Hastedt, S. (2012). ―More Money? The Impact of Larger Incentives on Response Rates in a Two Phase Mail Survey." Presentation at the Federal Committee on Statistical Methodology (FCSM) Research Conference. Washington, DC. January, 2012.

Nicoletti, C. and Perachhi, F. (2005). Survey Response and Survey Characteristics: Micro-level evidence from the European Commission Household Panel. Journal of the Royal Statistical Society: Series A (Statistics in Society) 168(4): 763-781. Stable URL: ftp://www.ceistorvergata.it/repec/rpaper/No-64-Nicoletti-Peracchi.pdf

Petrolia, D. and Bhattacharjee, S. (2009). Revisiting Incentive Effects: Evidence from a Random-Sample Mail Survey on Consumer Preferences for Fuel Ethanol. *Public Opinion Quarterly, 73(3): 537-550*.

Rendtel, U. and Harms, T. (2009). Weighting and Calibration for Households Panels. Methodology of Longitudinal Surveys. (ed. P. Lynn), Copyright 2006 (John Wiley & Sons, Ltd.).

Ryu, E., Couper M.P., Marans, R.W. (2006) Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error. *International Journal Public Opinion Research 18(1): 89-106* (first published online July 1, 2005) Stable URL: http://ijpor.oxfordjournals.org/content/18/1/89.short

Watson, N. and Wooden, M. (2009). Identifying Factors Affecting Longitudinal Survey Response. Methodology of Longitudinal Surveys. (ed. P. Lynn), Copyright 2006 (John Wiley & Sons, Ltd.). Stable URL: http://www.iser.essex.ac.uk/files/survey/ulsc/methodological-research/mols-2006/scientific-social-programme/papers/Watson.pdf

Willimack, D.K., Schuman, H., Pennell, B. and Lepkowskie, J.M. (1995). Effects of a Prepaid Nonmonetary Incentive on Response Rates and Response Quality in a Face-to-face Survey. Public Opinion Quarterly *59 (1): 78-92.* Stable URL: http://poq.oxfordjournals.org/content/59/1/78.short

Wobus, P. and de la Puente, M. (1995). Results from Telephone Debriefing Interviews: The Census Bureau's Spanish Forms Availability Test. Proceedings of the Section on Survey Research Methods 1995: 1040-1045. American Statistical Association. Alexandria, VA. Stable URL: http://www.amstat.org/sections/srms/proceedings/

Wright, M. (1995). The Effect of Pre-Notification on Mail Survey Response Rates: An Experimental Result. *Marketing Bulletin, 1995(6): 59-64, Research Note 3*.