Does it Pay to Send Multiple Pre-Paid Incentives? Evidence from a Randomized Experiment

Andrew C. Chang, Joanne W. Hsu, Eva Ma, Kate Bachtell, and Micah Sjoblom

2024-023

Does it Pay to Send Multiple Pre-Paid Incentives?
Evidence from a Randomized Experiment

April 15, 2024

Andrew C. Chang,[*] Board of Governors of the Federal Reserve System, andrew.c.chang@frb.gov
Joanne W. Hsu,[†] University of Michigan, jwhsu@umich.edu
Eva Ma, Board of Governors of the Federal Reserve System, eva.ma@frb.gov
Kate Bachtell, NORC at the University of Chicago, bachtell-kate@norc.org
Micah Sjoblom, NORC at the University of Chicago, sjoblom-micah@norc.org

Abstract

To encourage survey participation and improve sample representativeness, the Survey of Consumer Finances (SCF) offers an unconditional pre-paid monetary incentive and separate post-paid incentive upon survey completion. We conducted a pre-registered between-subject randomized control experiment within the 2022 SCF, with at least 1,200 households per experimental group, to examine whether changing the pre-paid incentive structure affects survey outcomes. We assess the effects of: (1) altering the total dollar value of the pre-paid incentive ("incentive effect"), (2) giving two identical pre-paid incentives holding the total dollar value fixed ("reminder effect"), and (3) offering multiple pre-paid incentives of different amounts holding the total dollar value fixed ("slope effect") on survey response rates, interviewer burden, and data quality. Our evidence indicates that a single $15 pre-paid incentive increases response rates and maintains similar levels of interviewer burden and data quality, relative to a single $5 pre-paid incentive. Splitting the $15 into two pre-paid incentives of different amounts increases interviewer burden though lengthening time in the field without improving response rates, reducing the number of contact attempts needed for a response, or improving data quality, regardless of whether the first pre-paid is larger or smaller than the second.

Keywords: pre-paid incentives; unconditional incentives; sequential incentives; response rates; surveys; data quality; household finance
JEL Codes: C83; C93; G5

Introduction

An increasingly challenging environment for recruiting participants for surveys poses many risks for researchers who need to balance achieving representative samples and maintaining data quality with controlling financial costs and time in the field. Incentives, including pre-paid incentives, which are unconditional on survey completion, are one important tool to encourage survey response. While surveys typically employ a single pre-paid incentive, during data collection pre-paid incentives can be structured in a variety of ways.

The Survey of Consumer Finances (SCF),[1] sponsored by the Federal Reserve Board (FRB), began employing pre-notification postcards followed by a pre-paid incentive of five dollars cash with invitation letters for the 2016 wave, based on findings from Hsu, Schmeiser, Haggerty, and Nelson (2017). To explore how changes in the structure of respondent incentives could improve response rates, duration in the field, and data quality we embedded an 8 week field experiment within the 2022 SCF. We randomly assigned respondents across 6 groups with varying pre-paid incentive structures and amounts, including one group with the previously used single pre-paid incentive of $5. In our experiment, we tested three different conditions: (1) altering the total dollar value of the pre-paid incentive ("incentive effect"), (2) giving two identical pre-paid incentives holding the total dollar value fixed ("reminder effect"), and (3) offering multiple pre-paid incentives of different amounts holding the total dollar value fixed ("slope effect").

We analyze the experimental results to determine the effects of different pre-paid incentive structures on survey response rates, interviewer burden, and data quality using a mixed-mode (phone and face-to-face) survey on household finances. The results provide insights into the costs and benefits of different designs and values of pre-paid incentives for completing an in-depth interviewer-administered survey on a sensitive topic. Moreover, by offering incentive amounts in a variety of values, we investigate whether the relationship between higher incentive amounts and completion rates is monotonic or deteriorates at larger values.

We find that larger pre-paid incentives yielded higher response rates of about 2 or 3 percentage points (around a third of the baseline response rate) with no deleterious effect on other survey outcomes, and that it is better to send the incentive as a single payment, rather than splitting it into multiple payments. Splitting the incentive into payments of different amounts increased field interviewer burden through increasing time in the field without improving response rates, reducing the number of contact attempts needed for a response, or improving data quality.

Theoretical Background on the Response to Pre-paid Monetary Incentives

An extensive body of empirical and theoretical research supports the use of incentives as part of a broader strategy to promote survey completion. An abundance of research has shown that unconditional pre-paid incentives are particularly effective relative to conditional post-paid

---

[1] Board of Governors of the Federal Reserve System (2023b). The SCF collects information on US household income, wealth, debts, and other financial outcomes. See Aladangady, Bricker, Chang, Goodman, Krimmel, Moore, Reber, Henriques Volz, and Windle (2023) for a description of the SCF.

incentives (Blohm and Koch 2021). In accordance with theories of social exchange, noncontingent pre-paid incentives can encourage survey response by providing a psychological sense of obligation to return the favor of the incentive (Gouldner 1960; Dillman 1978). Leverage-saliency theory provides another mechanism: a pre-paid incentive could help establish trust that a respondent will honor the request and complete the survey (Groves, Singer, and Corning 2000). That said, incentives could appeal to external motivations, which are typically less effective than intrinsic or altruistic motivations at generating compliance with a survey request (Hansen 1980).

The effectiveness of incentives varies by design and size. Different incentive amounts may generate a varying leverage or stronger feelings of obligation. According to theories of economic exchange, respondents respond to surveys when the overall benefits outweigh the costs, and thus larger incentives should yield higher response rates (Biner and Kidd 1994). However, response rates may not increase monotonically with the size of the incentive. Some studies have found that the relationship between the size of pre-paid incentives and response rates is nonlinear (Warriner, Goyder, Gjertsen, Hohner, and McSpurren 1996; Trussell and Lavrakas 2004).[2] One possible explanation is that the size of the incentive may appeal differentially to factors of low or high leverage, feelings of reciprocity, or perceptions of survey legitimacy. In addition, excessively large incentives could lead respondents to distrust the survey or appeal too heavily to extrinsic motivators to be effective. That said, an analysis of an increase in the SCF's post-paid incentive from 2007 to 2010 found that the increase reduced the contact attempts and time in the field needed for a response while maintaining data quality (Bricker 2014). And a 2014 experiment that imitated the instrument and field strategy of the SCF found no negative effects for very large conditional post-paid incentives (Hsu, Schmeiser, Haggerty, and Nelson 2017).

Most surveys employ pre-paid incentives once during data collection.[3] But given the proliferation of junk mail, individuals may pay less attention to their postal mail or may miss an initial mailer or incentive entirely. Consequently, some studies have investigated the use of repeated (sequential) pre-paid incentives with the hope that a repeat incentive will elicit more careful (or any) reading and consideration of the survey materials by the respondent. Messer and Dillman (2011) found that following up an initial request that included a $5 pre-paid incentive with a second $5 pre-paid incentive via priority mail to nonresponders increased the response rate of a state-wide web survey from 59% to 68%. In the context of a different mail-web survey, Wagner, West, Couper, Zhang, Gatward, Nishimura, and Saw (2023) found that following up a $2 pre-paid incentive included in the initial mailing with an additional $5 sent via priority mail considerably increased response rates. However, in both of these studies the effect of the second incentive cannot be disentangled from the effect of the priority mailing.

---

[2] Relatedly, Han, Montaquila, and Brick (2013, Table 2) find that the size of a pre-paid incentive affects response rates conditional on how quickly eligible respondents complete a screener.

[3] Correspondingly, there are a large number of studies that evaluate whether a single pre-paid incentive affects response rates. For example: Hsu, Schmeiser, Haggerty, and Nelson (2017); Frederiks, Romanach, Berry, and Toscas (2020); Jackson, McPhee, and Lavrakas (2020); Powell, Geronimo-Hara, Tobin, Donoho, Sheppard, Walstrom, Rull, and Faix (2023).

Dykema, Stevenson, Assad, Kniss, and Taylor (2021) conducted an experiment in a mail survey of physicians and found that "second incentives [sent a month after the first incentives] were associated with higher response rates and lower costs per completed survey" but no measurable effect on item nonresponse. Dillman, Smyth, and Christian (2014, p. 424) now recommend researchers in mixed-mode studies to include a second cash incentive with their follow-up communications to provide opportunities for "later communications [to] be read, and hopefully acted upon, thereby increasing overall response." That said, Dykema, Jaques, Cyffka, Assad, Hammers, Elver, Malecki, and Stevenson (2015) found that a second pre-paid incentive, again targeting nonresponders, did not increase response rates.

Few studies have investigated the differential effect of repeated incentives that are increasing or decreasing in size, what we call the "slope effect". One example is Dykema, Stevenson, Assad, Kniss, and Taylor (2021), who found few measurable differences in response rates or item response for second incentives that are larger than the first incentive, relative to second incentives that are smaller than the first. Similarly, they found the representativity of responders for either of these conditions was not significantly different from benchmark administrative data, so the slope effect in their study appears to have made no differences in survey outcomes.

In our study, we analyze pre-paid incentives in the context of a nationally representative mixed mode (face-to-face and phone) survey in which respondents are initially contacted via mail. While most existing studies involved sending second incentives only to nonresponders, our study delivered second mailings and incentives to potential respondents two weeks after the first, regardless of response.

Methods

Overview of the Survey of Consumer Finances

The SCF is a nationally representative survey on the finances of US households, conducted on a different cross-section of US families triennially (Board of Governors of the Federal Reserve System, 2023b). Topics covered include income, assets, debts, other financial characteristics, and economic behavior. The survey is administered by field interviewers (FIs) and, historically, is primarily conducted face-to-face.

Given the high concentration of wealth in the United States, the SCF uses a dual-frame sample to ensure coverage across the full distribution of wealth. The SCF employs both an address-based multistage nationally representative area-probability (AP) sample complemented by a stratified list sample specifically designed to oversample wealthy Americans.[4] In the face of an increasingly challenging environment for survey response rates over the past two decades, the SCF has repeatedly extended time and added expenses in the field. Each wave since 2004 required an average extension of 2.5 additional months beyond the target field period of eight months, with the 2019 and 2022 waves needing extensions of about 4 months.[5]

---

[4] See Kennickell (2005) for a discussion of the sampling procedure.

[5] In 2022 the AP sample response rate was about 42 percent, and the list sample response rate was about 27 percent, using RR1 from AAPOR (2015). See the appendixes to Bhutta, Bricker, Chang, Dettling, Goodman, Hsu, Moore, Reber, Henriques Volz, and Windle (2020); Aladangady, Bricker, Chang, Goodman, Krimmel, Moore, Reber,

Study Design

The 2022 SCF, which was sponsored by the FRB with cooperation from the Statistics of Income Division (SOI) at the Internal Revenue Service, and conducted by the National Opinion Research Center (NORC) at the University of Chicago, included a pre-registered between-subject randomized control experiment within the AP sample with six different pre-paid incentive groups (a 1x6 cell design).[6] We refer to particular groups using the format $[First]/$[Second Pre-Paid Amount]. One group ($5/$0), structured to be identical to the pre-paid incentive from the 2016 and 2019 SCFs, received a single $5 pre-paid incentive, followed by a second mailer with no monetary incentive. The other five treatment groups, shown in Table 1, facilitated testing the effects of: (1) larger total incentive payments (the incentive effect), (2) two incentive payments per respondent against a single pre-paid incentive with the same total value (the reminder effect), and (3) second pre-paid incentives that are larger or smaller in size than the first incentive, controlling for total value (the slope effects).

Table 1: Treatment Group Definitions

Group	First Pre-paid incentive amount	Second Pre-paid incentive amount	Sample Size	Notes
$5/$0	$5	-	2,152	The 2016/2019 SCF incentive, control for total $ amount for single pre-paid incentives
$5/$5	$5	$5	1,292	Baseline multiple mailer treatment
$10/$0	$10	-	1,291	Total $ amount control for baseline multiple mailers ($5/$5 group)
$5/$10	$5	$10	1,293	Tests upward slope of incentive
$10/$5	$10	$5	1,291	Test downward slope of incentive
$15/$0	$15	-	1,297	Total $ amount control for slope conditions ($5/$10 and $10/$5 groups)

We stratified our randomization by National Frame Area (NFA, a primary sampling unit of geography used by NORC to create the AP frame) across all NFAs in the AP sample. Randomization was conducted by NORC with a quasi-random number generator with no rerandomization. Due to limits on drawing the sampling frame for the 2022 SCF and the simultaneous nature of the treatment across all sample units, households were not reallocated across treatment groups to balance treatment group sizes when a household was out of scope.

We calibrated experimental group sample sizes based on the treatment effect sizes from Hsu, Schmeiser, Haggerty, and Nelson (2017), which is the closest paper to ours in terms of institutional setting. We selected sample sizes to give 80% power for our expected treatment

---

Henriques Volz, and Windle (2023) for information on the response rates and field period duration. In the SCF, households must answer all critical questions within the survey instrument for an interview to be complete. The 2022 SCF codebook denotes which questions are critical (Board of Governors of the Federal Reserve System 2023a). There is no SCF standard for partial completes.

[6] Our pre-registration plan is on the Open Science Framework under "2022 SCF Pre-paid Incentives Experiment" at https://dx.doi.org/10.17605/OSF.IO/BXJNE (Chang 2023). The pre-registration plan also includes a plan for analysis of the experimental data.

effect on response rates.[7] The resulting sample sizes, shown in Table 1, are at least 1,200 households per treatment group and are over eight times larger than those featured in Hsu, Schmeiser, Haggerty, and Nelson's work. Similarly, the treatment group-level sample sizes are two to four times larger than Messer and Dillman (2011) and Dykema et al. (2015, 2021), who also look at the effects of second pre-paid incentives on response rates. Our larger sample sizes give us much more statistical power to detect treatment effects.

Fielding of our experiment began by mailing all households in the AP sample a postcard introducing the SCF between late March and early April 2022. The first envelope mailing, sent on April 6th, 2022 with the USPS, included an invitation letter asking the head of household to provide their contact information through a secure website and was accompanied by the first pre-paid cash incentive—$5, $10, or $15, depending on the group.[8] A second envelope, mailed two weeks later with the USPS, included a pre-contact letter indicating that a field interviewer would be reaching out to describe the study further and schedule an interview. The second envelope was accompanied by a second pre-paid cash incentive equal to $5 or $10 for three of the six experimental groups. Households in the remaining three groups did not receive cash with the second envelope. All households received a second envelope, so the amount of the second pre-paid cash incentive was not dependent on survey completion. The invitation letter and all additional materials left by the interviewer included a toll-free number that a respondent could call with questions at any time during the experiment or to complete an interview.

Envelopes included transparent windows, so households could see from an examination of the exterior that there was cash inside. See Figure 1 for a mock-up of the mailing envelope. Envelopes with $5 had a single $5 bill in the window. Envelopes with $10 or $15 had a $10 bill placed in the window, and those that received $15 also had a $5 bill inside the envelope, placed directly behind the $10 bill, so the $5 bill was not visible from the exterior. Therefore, the total amount of the pre-paid cash incentive was only discernable from opening the envelope. Other visual elements were minimized to distinguish the envelope from commercial sales materials and avoid detracting from the cash enclosure.

---

[7] See our pre-registration plan for additional details on the power calculations (Chang 2023).
[8] Households who provided their contact information through the secure website were paid an additional $10, an amount that was not dependent on the household's experimental group.

Figure 1: Mailing Envelope Mock-Up

Front:
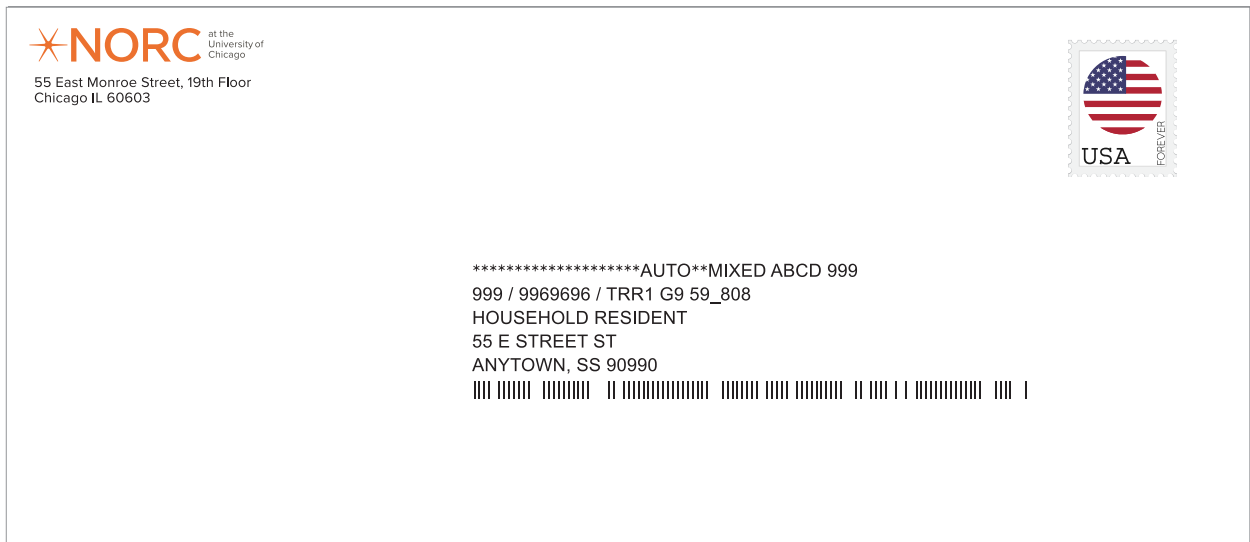
Back:
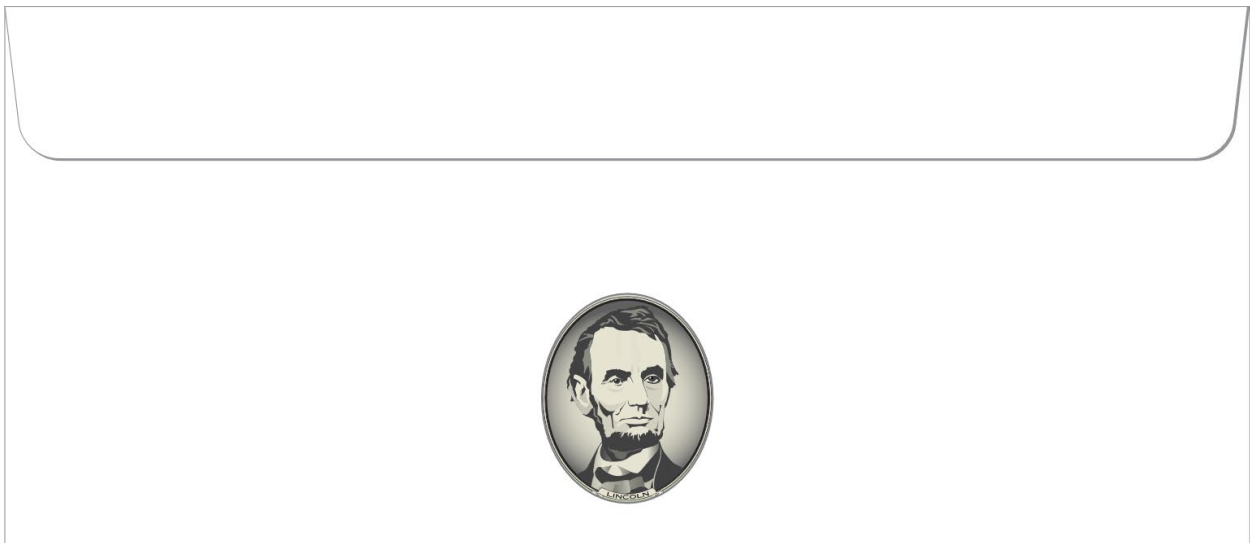
Interviewer outreach to sampled households began in mid-April 2022. Initial contact attempts focused on households that had provided contact information through the secure website. To avoid experimenter effects, field staff were not aware of a household's pre-paid incentive amount(s), though this information could be voluntarily given to field staff by the household after initial contact. The field period for our experiment ended on June 1st, 2022, 8 weeks after we mailed the first envelope. Completed interviews followed the same post-paid incentive structure regardless of pre-paid treatment assignment.

Model

We used pairwise comparisons with regressions/linear probability models of the form:

$$y_{i,s} = \sum_{\forall s} \alpha_{s,y} + \beta_y \, treatment_{i,s} + \epsilon_{i,s,y} \quad (1)$$

We estimated the models with ordinary least squares with the omitted category as the control group. The outcome y for individual i in stratum (NFA) s is a function of a full vector of stratum dummies ($\alpha_{s,y}$), following the recommendation of Bruhn and McKenzie (2009), and a treatment group indicator. The treatment indicator depended on the hypothesis being tested. We weighted analyses by the NFA's inverse probability of selection into the 2022 SCF. Standard errors were calculated using Huber-White heteroskedasticity consistent standard errors (White, 1980).

Research Questions and Analysis Methods

Our experimental design facilitated testing the following hypotheses:

Hypothesis 1: Larger values of total cash pre-paid incentives affect survey outcomes (the incentive effect).

- *Hypothesis 1a: $5 vs. $10.* Include $10/$0 and $5/$0 groups in equation (1), test β_y (treatment group is $10/$0).
- *Hypothesis 1b: $10 vs. $15.* Include groups $15/$0 and $10/$0 in equation (1), test β_y (treatment group is $15/$0).
- *Hypothesis 1c: $5 vs. $15.* Include groups $15/$0 and $5/$0 in equation (1), test β_y (treatment group is $15/$0).

Hypothesis 2: A second pre-paid incentive (through a reminder effect), controlling for the total pre-paid incentive cash amount across envelopes (the incentive effect), affects survey outcomes under no change in the incentive amount between incentives (no slope effect).

- Include $5/$5 and $10/$0 groups in equation (1), test β_y (treatment group is $5/$5).

Hypothesis 3: The increasing or decreasing the amount of pre-paid incentive cash between multiple envelopes (the slope effect), controlling for the total cash amount (the incentive effect) affects survey outcomes.

- *Hypothesis 3a: Increasing slope of incentive.* Include $5/$10 and $15/$0 groups in equation (1), test β_y (treatment group is $5/$10).
- *Hypothesis 3b: Decreasing slope of incentive.* Include groups $10/$5 and $15/$0 in equation (1), test β_y (treatment group is $10/$5).

We tested survey outcomes related to response rates, interviewer burden, and data quality. For response rates, we employed AAPOR (2015)'s RR1 with the SCF standard for completed interviews, and also looked at the share of invitees who scheduled interview appointments. For interviewer burden, we analyzed: the number of contact attempts needed to complete an interview; the number of contact attempts needed to elicit a website visit or a scheduled

appointment;[9] and the duration of time between when the first pre-paid was mailed and a completed interview. Finally, for data quality, we looked at two outcomes. The first is item response rates: the share of questions that the respondent completed from the instrument of those that the respondent was eligible for. The second data quality outcome is the item response rate for questions that required a dollar value as a response that were answered with an exact dollar value, among those interviews completed with at least 50% of eligible dollar value questions being non-missing.[10] Table 2 displays sample sizes and average outcomes for each group.[11]

---

[9] We had initially specified adding contact attempts before a callback to this measure, but for operational reasons we were unable to effectively track callbacks.

[10] Affecting the survey response rate through our pre-paid incentive treatment may also have affected the participation of respondents who are differentially likely to complete the survey with missing answers. See the Appendix for details on how we used SCF paradata to construct these measures. See Meyer, Mok, and Sullivan (2015) for a review of survey data quality.

[11] Our response rates are not comparable to the overall SCF response rates because all experimental outcomes were measured within 8 weeks of the start of the experiment, in contrast to an approximately year-long field period for the entire SCF.

Table 2: Sample Sizes, Response Rates, Interviewer Burden, and Data Quality by Experimental Group

	Experimental Group: Amount of First/Second Pre-paid Incentive						
	$5/$0	$5/$5	$10/$0	$5/$10	$10/$5	$15/$0	Total
N	2,152	1,292	1,291	1,293	1,291	1,297	8,616
Worked	1,411	865	849	844	873	870	5,712
% Worked	65.6	67.0	65.8	65.3	67.6	67.1	66.3
Response Rates							
Appointments (count)	270	186	200	193	201	193	1,243
% Appointments	12.5	14.4	15.5	14.9	15.6	14.9	14.4
Completes (count)	140	102	111	115	110	121	699
% Completes	6.5	7.9	8.6	8.9	8.5	9.3	8.1
Interviewer Burden							
Avg Attempts Before Appt	1.6	1.8	1.5	1.5	1.6	1.4	1.6
Standard Deviation	1.6	1.9	1.6	1.5	1.7	1.9	1.7
Avg Attempts Before Complete	3.2	3.7	3.4	3.5	3.6	3.3	3.4
Standard Deviation	1.8	2.1	1.9	1.7	2.2	1.8	1.9
Avg Days Before Complete	29.2	26.3	29.4	31.0	30.1	26.5	28.8
Standard Deviation	13.8	13.2	14.0	13.8	13.4	13.6	13.7
Data Quality							
Avg % Questions Answered	97.9	97.6	98.4	97.8	97.5	98.0	97.9
Standard Deviation	2.2	2.6	1.9	2.6	2.8	2.0	2.4
Avg % Dollar Questions Answered	86.8	85.6	86.4	88.0	86.2	89.2	87.1
Standard Deviation	14.5	15.6	17.9	17.2	15.4	17.6	16.3

Note: Averages are weighted by the NFA's inverse probability of selection. Worked implies that a household received attention from field interviewers or managers. "Avg Attempts Before Appt" includes contact attempts before a respondent scheduled an appointment or made a website visit.

We corrected for multiple hypothesis testing by controlling for the family-wise error rate (FWER) using the Westfall-Young (1993) free-step down procedure, implemented by Jones, Molitor, and Reif (2018). We defined families for FWER corrections as outcomes for response rates, interviewer burden, and data quality grouped for each hypothesis (1a, 1b, etc.). We used ten thousand replications per bootstrap resample.

All of our statistical tests were two-sided for the null of no effect vs. the alternative of an effect. In the discussion that follows, we report FWER-corrected p-values, which tend to be more conservative (larger) than the p-values that are not adjusted for multiple hypothesis testing.

Results

Incentive effects

Table 3 displays our results for incentive effects, comparing experimental groups that received only one pre-paid incentive of $5, $10, or $15. A $10 pre-paid incentive increased the share of invitees who made an appointment for an interview by 3.1 p.p. (22.5%, p=0.04), and increased the share of completed interviews by 2.2 p.p (30.1%, p=0.04). We are not able to statistically distinguish different rates of appointments or completions between those receiving $10 and $15 pre-paid incentives. Increasing the pre-paid incentive from $5 to $15, however, did increase the response rate by 2.9 p.p. (39.9%, p=0.02). There is an increase in both the magnitude and precision of the estimated effect of the $15/$0 treatment on interview completions, relative to the $10/$0 treatment. Our results suggest that, at least at the level of increasing the pre-paid incentive from $5 to $15, that increasing pre-paid incentives do not appear to have unintended negative consequences on respondent cooperation.

There is no statistically significant difference in interviewer effort between the treatment groups; time in the field and number of contact attempts before an appointment or a complete were the same across groups. Similarly, there is no effect on data quality in the pairwise comparisons across the three pre-paid incentive values.

Table 3: Incentive Effect

b/se/p/N	$5/$0 vs. $10/$0	$10/$0 vs. $15/$0	$5/$0 vs. $15/$0
Response Rates			
(1) Appointment	3.10**	-1.22	1.83
	(1.40)	(1.60)	(1.38)
	[0.04]	[0.61]	[0.18]
	3,443	2,588	3,449
(2) Complete	2.16**	0.72	2.87**
	(1.09)	(1.30)	(1.12)
	[0.04]	[0.61]	[0.02]
	3,443	2,588	3,449
FI Burden			
(3) Attempts Before First Appt	-0.10	-0.10	-0.09
	(0.18)	(0.21)	(0.21)
	[0.90]	[0.77]	[0.86]
	514	424	506
(4) Attempts Before Complete	0.13	-0.20	0.13
	(0.33)	(0.34)	(0.31)
	[0.90]	[0.77]	[0.86]
	251	232	261
(5) Days Before Complete	0.98	-2.76	-3.57
	(2.23)	(2.56)	(2.22)
	[0.90]	[0.55]	[0.24]
	251	232	261
Data Quality			
(6) Pct. Answers	0.37	-0.56	0.20
	(0.38)	(0.37)	(0.35)
	[0.49]	[0.21]	[0.60]
	251	232	261
(7) Pct. Dollar Values	-2.80	3.25	1.97
	(2.79)	(2.83)	(2.38)
	[0.49]	[0.23]	[0.60]
	250	230	260

Description: Table displays treatment coefficients (treatment effect sizes), standard errors in parentheses, FWER-corrected p-values in brackets, and the regression's sample size below each p-value from the regression of the respective outcome on the treatment indicator and stratum dummies. Each numbered row is a different outcome indicated by the count of: (1) appointments, (2) completes, (3) field interviewer (FI) attempts before a scheduled appointment or website visit, (4) FI attempts before a complete, (5) days in the field before a complete, followed by (6) the item response rate (%), and (7) the item response rate for dollar value questions considering only responses with a point estimate (%). Hypothesis tests are two sided against the null of no effect. *p<0.1, **p<0.05, ***p<0.01

Interpretation: A higher total incentive amount increases appointments and completed interviews. There is no effect of the total incentive amount on FI burden or data quality, so higher total incentive amounts do not appear to have unintended negative consequences for survey outcomes.

Reminder effects

Table 4 displays results for comparing the group that received two $5 pre-paid incentives ($5/$5) to the group that received $10 in a single envelope ($10/$0). There are no statistically significant differences in response rates, interviewer effort, or data quality across the two groups.

Table 4: Reminder Effect

b/se/p/N	$5/$5 vs. $10/$0
Response Rates	
(1) Appointment	-1.17
	(1.59)
	[0.64]
	2,583
(2) Complete	-0.90
	(1.25)
	[0.64]
	2,583
FI Burden	
(3) Attempts Before First Appt	0.23
	(0.22)
	[0.31]
	416
(4) Attempts Before Complete	0.58
	(0.43)
	[0.31]
	213
(5) Days Before Complete	-3.61
	(2.57)
	[0.31]
	213
Data Quality	
(6) Pct. Answers	-0.62
	(0.54)
	[0.38]
	213
(7) Pct. Dollar Values	-1.58
	(3.48)
	[0.63]
	212

Description: Table displays treatment coefficients (treatment effect sizes), standard errors in parentheses, FWER-corrected p-values in brackets, and the regression's sample size below each p-value from the regression of the respective outcome on the treatment indicator and stratum dummies. Each numbered row is a different outcome indicated by the count of: (1) appointments, (2) completes, (3) field interviewer (FI) attempts before a scheduled appointment or website visit, (4) FI attempts before a complete, (5) days in the field before a complete, followed by (6) the item response rate (%), and (7) the item response rate for dollar value questions considering only responses with a point estimate (%). Hypothesis tests are two sided against the null of no effect. *p<0.1, **p<0.05, ***p<0.01

Interpretation: Controlling for the total incentive amount, splitting the pre-paid incentive into two mailers of the same amount (the reminder effect) has no effect on response rates, FI burden, and data quality.

Slope effects

Table 5 presents results for the hypotheses involving second incentives that are larger (or smaller) than the first incentive. The comparison group received the same total amount in a single incentive. There are no measurable pairwise differences in response rates or data quality. However, the positive slope group ($5/$10) took about 5 additional days to complete the interview (18.8%, p=0.09), though there are no differences in the number of contact attempts. Similarly, the negative slope group ($10/$5) also took almost 5 days longer to respond (p=0.09). Therefore, splitting the same total incentive amount into two separate envelopes of different amounts, on net, worsened survey outcomes.

Table 5: Slope Effect

b/se/p/N	$5/$10 vs. $15/$0	$10/$5 vs. $15/$0
Response Rates		
(1) Appointment	0.55	0.70
	(1.60)	(1.58)
	[0.84]	[0.82]
	2,590	2,588
(2) Complete	-0.58	-0.60
	(1.31)	(1.31)
	[0.84]	[0.82]
	2,590	2,588
FI Burden		
(3) Attempts Before First Appt	0.16	0.13
	(0.20)	(0.22)
	[0.63]	[0.53]
	413	435
(4) Attempts Before Complete	0.13	0.44
	(0.32)	(0.38)
	[0.65]	[0.39]
	236	231
(5) Days Before Complete	4.95*	4.98*
	(2.37)	(2.45)
	[0.09]	[0.09]
	236	231
Data Quality		
(6) Pct. Answers	-0.19	-0.31
	(0.41)	(0.49)
	[0.61]	[0.73]
	236	231
(7) Pct. Dollar Values	-2.42	-0.37
	(2.66)	(3.35)
	[0.54]	[0.91]
	235	230

Description: Table displays treatment coefficients (treatment effect sizes), standard errors in parentheses, FWER-corrected p-values in brackets, and the regression's sample size below each p-value from the regression of the respective outcome on the treatment indicator and stratum dummies. Each numbered row is a different outcome indicated by the count of: (1) appointments, (2) completes, (3) field interviewer (FI) attempts before a scheduled appointment or website visit, (4) FI attempts before a complete, (5) days in the field before a complete, followed by (6) the item response rate (%), and (7) the item response rate for dollar value questions considering only responses with a point estimate (%). Hypothesis tests are two sided against the null of no effect. *p<0.1, **p<0.05, ***p<0.01

Interpretation: Controlling for the total incentive amount, splitting the pre-paid incentive into two mailers of different amounts increases time in the field, regardless of whether the second payment is higher or lower than the first (the slope effect worsens FI burden).

Discussion

In our experiment, the largest results were concentrated in the incentive effect. In the ranges that we studied, larger pre-paid incentives yielded much higher response rates. We found no deleterious effect of larger pre-paid incentives on response rates. Larger pre-paid incentives increased response rates by about 2 to 3 percentage points, about a third of the base response rate of 7%.

When holding the total incentive amount constant, response rates did not measurably improve when splitting the total incentive into two mailings. The relative amounts of the two mailings also did not matter for response rates.

Overall, there were very few measurable differences in interviewer effort and data quality across experimental groups. The one exception was the two slope conditions, $5/$10 and $10/$5, which both required more days in the field for survey completion than the $15/0 group. It is possible that the slope conditions set expectations by households that they would receive additional mailers containing pre-paid incentives, so households may have delayed responding to the survey as a result. As a longer field period places more burden on the field interviewers, for example requiring interviewers to keep track of a larger portfolio of potential respondents, our findings suggest that splitting an incentive between two mailers with different amounts can worsen survey outcomes. Furthermore, the remaining cases late in the field period are likely to have lower response propensities, placing an even greater burden on field interviewers. Our results imply that the total incentive amount, delivered early in the field period, may be more important to individuals than the number of times they receive an incentive.

In contrast to previous literature in which second incentives were sent a lengthy period after the first, in our experiment the two incentives were sent two weeks apart. Our results may have differed if our second mailings were sent with a greater delay, in which case respondents may not mentally connect the two incentives with each other as strongly as they did in our experiment. Alternatively, delaying the second incentive could give the first incentive more opportunity to take effect, and could better facilitate targeting cases that interviewers had already attempted and failed to complete. While targeting nonresponders could reduce the total amount spent on incentives it would also be operationally more difficult in many settings, particularly when the arrival of the pre-paid incentive needs to be coordinated with other field strategies, like the availability of an interviewer to work a case. In addition, second incentives in the dollar range we considered might not make a difference to highly reluctant respondents, so the benefit of sending second pre-paid incentives to all potential respondents (reduced operational difficulty) may outweigh the drawback (higher survey budget for incentives).

References

Aladangady, Aditya, Jesse Bricker, Andrew C. Chang, Sarena Goodman, Jacob Krimmel, Kevin B. Moore, Sarah Reber, Alice Henriques Volz, and Richard A. Windle. 2023. "Changes in U.S. Family Finances from 2019 to 2022: Evidence from the Survey of Consumer

Finances." Washington: Board of Governors of the Federal Reserve System, October, doi: https://doi.org/10.17016/8799

American Association for Public Opinion Research (AAPOR). 2015. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Lenexa, KS.

Bhutta, Neil, Jesse Bricker, Andrew C. Chang, Lisa J. Dettling, Sarena Goodman, Joanne W. Hsu, Kevin B. Moore, Sarah Reber, Alice Henriques Volz, and Richard A. Windle. 2020. "Changes in U.S. Family Finances from 2016 to 2019: Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin* 106(5). doi: https://doi.org/10.17016/bulletin.2020.106

Biner, Paul M., and Heath J. Kidd. 1994. "The Interactive Effects of Monetary Incentive Justification and Questionnaire Length on Mail Survey Response Rates," *Psychology & Marketing* 11(5): 483–492.

Bledsoe, Ryan, and Gehard Fries. 2002. "Editing the 2001 Survey of Consumer Finances," In Proceedings of the Section on Survey Research Methods, 2002 Annual Meetings of the American Statistical Association, New York.

Blohm, Michael, and Achim Koch. 2021. "Monetary Incentives in Large-Scale Face-to-Face Surveys: Evidence from a Series of Experiments," *International Journal of Public Opinion Research* 33(3): 690-702.

Board of Governors of the Federal Reserve System. Division of Research and Statistics, Microeconomic Surveys. 2023a. "Codebook for 2022 Survey of Consumer Finances," http://web.archive.org/web/20231029104656/https://www.federalreserve.gov/econres/files/codebk2022.txt

Board of Governors of the Federal Reserve System. Division of Research and Statistics, Microeconomic Surveys. 2023b. "Survey of Consumer Finances," doi: https://doi.org/10.17016/datasets.001

Bricker, Jesse. 2014. "Survey Incentives, Survey Effort, and Survey Costs," Finance and Economics Discussion Series 2014-74. Washington: Board of Governors of the Federal Reserve System. doi: https://doi.org/10.17016/FEDS.2014.074

Bruhn, Miriam, and David McKenzie. 2009. "In pursuit of balance: Randomization in practice in development field experiments," *American Economic Journal: Applied Economics* 1(4): 200–232.

Chang, Andrew C. 2023. "2022 SCF Pre-Paid Incentives Experiment," Open Science Framework. Last updated April 14, 2022. doi: https://dx.doi.org/10.17605/OSF.IO/BXJNE.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley Interscience.

Dillman, Don A., Jolene D. Smyth, and Christian Leah Melani. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 4th ed. Hoboken, NJ: John Wiley & Sons

Dykema, Jennifer, John Stevenson, Nadia Assad, Chad Kniss, and Catherine A. Taylor. 2021. "Effects of Sequential Prepaid Incentives on Response Rates, Data Quality, Sample Representativeness, and Costs in a Mail Survey of Physicians," *Evaluation & the Health Professions* 4(3): 235-244.

Dykema, Jennifer, Karen Jaques, Kristen Cyffka, Nadia Assad, Rae Ganci Hammers, Kelly Elver, Kristen C. Malecki, and John Stevenson. 2015. "Effects of sequential prepaid incentives and envelope messaging in mail surveys," *Public Opinion Quarterly*, 79(4): 906–931.

Frederiks, Elisah R., Lygia M. Romanach, Adam Berry, and Peter Toscas. 2020. "Making energy surveys more impactful: Testing material and non-monetary response strategies," *Energy Research & Social Science* 63: 101409.

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration," *Public Opinion Quarterly* 64(3): 299–308. http://www.jstor.org/stable/3078721

Gouldner, Alvin W. 1960. "The Norm of Reciprocity: A Preliminary Statement," *American Sociological Review* 25(2): 161-178.

Han, Daifeng, Jill M. Montaquila, and J. Michael Brick. 2013. "Evaluation of Incentive Experiments in a Two-Phase Address-Based Sample Mail Survey," *Survey Research Methods* 7(3): 207-218.

Hansen, Robert A. 1980. "A Self-Perception Interpretation of the Effect of Monetary and Nonmonetary Incentives on Mail Survey Respondent Behavior," *Journal of Marketing Research* 17(1): 77–83.

Hsu, Joanne W., Maximilian Schmeiser, Catherine Haggerty, and Shannon Nelson. 2017. "The Effect of Large Monetary Incentives on Survey Completion: Evidence from a Randomized Experiment with the Survey of Consumer Finances," *Public Opinion Quarterly* 81(3): 736–747, doi: https://doi.org/10.1093/poq/nfx006

Jackson, Michael T., Cameron B. McPhee, and Paul J. Lavrakas. 2020. "Using Response Propensity Modeling to Allocate Noncontingent Incentives in an Address-Based Sample: Evidence from a National Experiment," *Journal of Survey of Statistics and Methodology* 8(2): 385-411.

Jones, Damon, David Molitor, and Julian Reif. 2018. "What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study," National Bureau of Economic Research Working Paper No. 24229.

Kennickell, Arthur B. 2005. "The Good Shepherd: Sample Design and Control for Wealth Measurement in the Survey of Consumer Finances," Paper presented at the Luxembourg Wealth Study Conference, Perugia, Italy. Available at http://www.federalreserve.gov/pubs/oss/oss2/papers/sampling.perugia05.2.pdf.

Messer, Benjamin L., and Don A. Dillman. 2011. "Surveying the general public over the internet using address-based sampling and mail contact procedures," *Public Opinion Quarterly*, 75(3): 429–457. doi: https://doi.org/10.1093/poq/nfr021.

Meyer, B.D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household surveys in crisis," *Journal of Economic Perspectives*, 29(4): 199-226.

Powell, Teresa M., Toni Rose Geronimo-Hara, Laura E. Tobin, Carrie J. Donoho, Beverly D. Sheppard, Jennifer L. Walstrom, Rudolph P. Rull, and Dennis J. Faix. 2023. "Pre-incentive Efficacy in Survey Response Rates in a Large Prospective Military Cohort," *Field Methods* 35(4): 392-408.

Trussell, Norm, and Paul J. Lavrakas. 2004. "The Influence of Incremental Increases in Token Cash Incentives on Mail Survey Response: Is There an Optimal Amount?" *Public Opinion Quarterly* 68(3): 349–367.

Wagner, James, Brady T. West, Mick P. Couper, Shiyu Zhang, Rebecca Gatward, Raphael Nishimura, and Htay-Wah Saw. 2023. "An Experimental Evaluation of Two Approaches for Improving Response to Household Screening Efforts in National Mail/Web Surveys," *Journal of Survey Statistics and Methodolog*y 11(1): 124–140. doi: https://doi.org/10.1093/jssam/smac024

Warriner, Keith, John Goyder, Heidi Gjertsen, Paula Hohner, and Kathleen McSpurren. 1996. "Charities, No; Lotteries, No; Cash, Yes Main Effects and Interactions in a Canadian Incentives Experiment," *Public Opinion Quarterly* 60(4): 542–562.

Westfall, Peter H., and S. Stanley Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279. John Wiley & Sons, 1993.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48(4): 817-838.

Appendix: Data Construction

Because we embedded the experiment within the 2022 SCF, our method of tracking experimental outcomes relied on paradata that is normally created with the fielding of the SCF, so the data are subject to certain operational constraints. We did not have a method in our pre-registration plan to address most of the issues with data construction in this appendix.

We used field interviewer Record of Calls (ROCs) to create the experimental outcomes for responses and interviewer effort (appointments, completes, contact attempts to schedule an appointment or website visit, contact attempts to complete, and calendar days to complete). What we observe in the ROCs is something like in Table A1:

Table A1: Example Record of Call (ROC)

Respondent ID #	Date (Month/Day)	ROC Type (Field Staff Action)
1	5/1	Voicemail
1	5/1	Voicemail
1	5/7	Appointment
1	5/14	Broken Appointment
1	5/15	Interview Complete
2	6/1	Out of Scope
3	4/10	Needs Review
3	4/12	Interview Complete
3	4/15	Appointment
4	5/1	Interview Complete
4	5/15	Out of Scope

Where each unique respondent ID # comes with a record entry date and a record type that indicates an action by field staff.

For a handful of interviews there are data discrepancies in the ROCs. Field staff actions listed in the ROCs may not appear on the date at which the actions actually occurred. This mismatch is due to technical issues that are independent of treatment status, such as synchronization issues between when actions are completed on field staff laptops and the NORC central servers, hard drive corruption of field staff laptops, or other miscellaneous technical issues. For example, for a particular respondent ID #, we can observe a respondent ID # with an appointment that appears on a date after the interview was completed (e.g., respondent ID #3 above), even though we know that an appointment should appear in the ROCs before an interview is complete. As a consequence of these technical issues, it is also possible to observe a particular respondent ID # with a series of field staff actions that should not be possible (e.g., a household deemed out of scope that is also recorded as a complete, see respondent ID #4 above).

Generally, these data discrepancies can be explained by a within-respondent swap of ROC entries (e.g., an appointment that was made on 4/10 is recorded on 4/15 for the correct respondent ID #) or a between-respondent swap of ROC entries (e.g., an appointment for respondent ID #3 gets mistakenly classified as respondent ID #4). Our assessment is that the technical issues causing these data discrepancies are more likely to cause a within-respondent swap than a between-respondent swap, and the data issues as a whole affect less than 1% of sampled respondent ROCs.

For our analysis we: (1) drop respondents (households) where we cannot explain the ROC history due to within-respondent ROC swapping (e.g., where both a refusal and a complete show up for the same respondent, there must have been a between-respondent swap of ROC entries to explain this pattern and we cannot tell whether the refusal or the complete is the correct ROC so we need to drop the respondent), (2) keep respondents where within-respondent swapping of ROCs can explain a time series of ROC entries that should not be plausible (e.g., an appointment that appears after a complete, the respondent potentially still made an appointment and completed the interview so we could still record both, even though appointments should show up before completes), and (3) take the ROC dates as-is for determining which actions happened during the pre-registered experimental period (i.e., ignoring potential effects of swaps, and associated measurement error, on the entry dates).

The technical issues only affect the outcome variables. Because the root causes of the ROC data discrepancies are technical issues that are independent of treatment status, the measurement error generated by these discrepancies is ignorable and, because less than 1% of interviews are affected, the increase in estimation noise from these discrepancies is negligible. At worst our hypothesis tests have p-values that are slightly too conservative (large) and are therefore slightly more likely to fail to reject the null hypothesis.

For creating the outcomes for contact attempts to schedule an appointment or website visit and contact attempts to complete, we counted multiple field staff actions on the same day to contact a household a single contact attempt. For example, interviewers occasionally attempt to call the same respondent several times in succession using several potential phone numbers. We treat the set of calls as a single good-faith effort to contact a respondent.

For creating our data quality measures, we used the original respondent-provided answers as entered into the computer-assisted personal interviewing (CAPI) software. We did not use any changes to the data by the SCF staff as a result of data review.[12]

---

[12] See Bledsoe and Fries (2002) for a description of the SCF's data review process.