

Cost Effective Mail Survey Design

Douglas Williams¹, W. Sherman Edwards¹, Pamela Giambo¹, Grace Kena²

¹Westat, 1600 Research Blvd. Rockville, MD 20850; ²Bureau of Justice Statistics, 810 Seventh St., NW
Washington, DC 20531

Proceedings of the 2018 Federal Committee on Statistical methodology (FCSM) Research Conference

Abstract

Policy needs for estimates at local levels have greatly increased, especially for data of uniform quality so that comparisons across areas and over time are valid. Currently, data collection by mail is the methodology best suited to meet these goals, although mixed Internet and mail methods are becoming more competitive. This paper describes experiments to explore cost-effective data collection methods that could be undertaken by local communities. The American Crime Victimization Survey (ACVS) is a prototype tested by the Bureau of Justice Statistics for local jurisdictions to get reliable information on victimizations and policing for comparisons to other areas and over time. The ACVS field test used a randomized block design to test the use of incentives (\$0, \$1, and \$2) and delivery methods (FedEx, regular mail) in a mail survey conducted in the 40 largest U.S. metropolitan areas. This paper describes the design and results of the experiments, conducted with a sample of more than 200,000 addresses. The treatments are compared with respect to response rates, achieved sample composition, and key outcome variables such as victimization and responses to community and policing items. In addition, we describe the relative cost per returned interview under each of the treatments. It is well known that incentives increase response rates in mail surveys, but this metric helps answer the question of which approach achieves the largest number of returned interviews for a fixed cost.

Key words: mail survey; response; incentives; data collection cost

1. Introduction

The desire to generate local or small-area estimates has enticed researchers for some time. Due cost of collecting data for local areas, statistical approaches have been explored (see Fay, Planty, and Diallo 2013, e.g.). These approaches are limited in their timeliness, requiring modeling data collected over several years, and the limited volume of data for small areas may affect the behavior of statistical models. Despite these challenges, the desire for quality and current data at a local level to inform policy decisions has increased in recent years.

One area where this desire has grown markedly is for information on criminal victimization, police performance, and attitudes toward police. One solution is for local areas or police jurisdictions to collect their own data rather than relying on a large national survey such as the National Crime Victimization Survey (NCVS). The NCVS currently reports national data on criminal victimization, but the scale of covering local areas across the nation is too costly. Alternatively, local areas lack the resources or infrastructure to field a complex in-person survey such as the NCVS.

Data collection by mail has become an attractive low-cost alternative to telephone or in-person, with well-established procedures. A mail data collection approach requires creating a validated mail questionnaire, and demonstrating a cost effective methodology. Williams, Brick, Edwards, and Giambo (in press), report the results of using mail questionnaires for estimates of victimization prevalence, showing strong correlations with the NCVS and achieving a relatively high response rate. This research has provided a mail questionnaire for local areas to collect information on victimization, leaving a need to demonstrate procedures that maximize response while showing responsible use of resources – that is, cost effectiveness.

There are a number of well-known procedures for improving response to mail surveys. The two most common are the use of incentives and special mail delivery (e.g., FedEx) for nonresponse follow-up. A large body of research shows that prepaid incentives in mail surveys are effective for improving response (Church 1993; Mercer, Caporaso, Cantor, and Townsend 2015; Singer and Ye 2013). The challenge lies in determining, then justifying the amount. While larger incentives are more effective, the effect is non-linear resulting in diminishing returns (Mercer et al., 2015). Additionally, larger amounts can result in rather dramatic increases in cost, especially for large samples.

Using a different or special method of delivery has long been used as a method to increase mail response (Dillman, Christenson, Carpenter, and Brooks 1974; Heberlein and Baumgartner 1978). Dillman and his colleagues used certified mail in the last mail nonresponse follow-up, but this requires a named recipient. Hager, Wilson, Pollak, and Rooney (2003), found that the use of FedEx over standard mail improved response in an initial mailing to nonprofit organizations. Special delivery options are much more expensive than, for example, first class standard mail. Even when used in a final contact for nonresponse follow-up, if overall response is low the volume of special mailings can be high, increasing cost.

It is difficult to consider the use of prepaid incentives or special mailing methods as cost effective. These are usually used to maximize survey response to efficiently utilize a sample, or in the hopes of reducing biases by encouraging response from groups with low response propensities. Often incentives are considered too expensive due to the large upfront cost, especially with large samples, or that too many would be discarded in unopened survey packages. This is a shortsighted view. It either relies on unrealistic response rate assumptions without efforts like incentives to improve response, or fails to consider the value of the increase in response. In other words, it is often cheaper to accept less. There is limited research available on how incentives affect data collection costs. Montaquilla, Brick, Williams, Kim, and Han (2013) observed that the cost per returned interview of a \$5 prepaid incentive was the same as no incentive. Biemer, Murphy, Zimmer, Berry, Deng, and Lewis (2018) examined the cost ratios of different incentive amounts in an address-based sample using mail and Web. They found that even large promised incentives of \$10 or \$20 combined with prepaid incentives of \$5 had small effects on cost. These limited findings suggest that incentives may be important in moderating or possibly reducing data collection cost.

In this paper, we review the results of an experiment looking at the use of survey enhancements, specifically, prepaid monetary incentives and FedEx delivery for our final nonresponse contact. We start with detailed results on the separate effects of different incentive levels and FedEx delivery for nonresponse follow-up. This provides information on the change each survey enhancement has on response (i.e., yield). Our review also includes combined results when these enhancements are used together. This addresses the question of whether any positive effects (from both incentives and FedEx delivery) are cumulative or whether using one reduces the effectiveness of both. We then examine the cost effectiveness of each, presenting cost in terms of yield and change in cost for a fixed yield. Our yield measure is cost per complete which factors the total cost across all survey returns. In the case of survey enhancements, this is a “return on investment” measure as increases in response would lower the cost per complete. Changes in cost for a fixed yield, show how much more, or less, it would cost to reach the same yield (number of completes) without survey enhancements compared to the yields for each survey enhancement. That is, what is the change in cost to increase the sample to reach the same number of returned surveys achieved with incentives or FedEx delivery?

2. Data and Methods

2.1. American Crime Victimization Survey

The American Crime Victimization Survey (ACVS) was designed as a companion survey to the National Crime Victimization Survey (NCVS). Like the NCVS, the ACVS was sponsored by the Bureau of Justice Statistics. The ACVS was developed to test the feasibility of collecting general information on victimization prevalence that could be administered at the local area (e.g., small cities) or by police jurisdictions. While both the NCVS and ACVS collect information on victimization, there are a number of differences between the two surveys.

An important objective of the ACVS is low operational burden, meaning that it must not require an extensive survey operations infrastructure to support and administer. The NCVS is a large-scale computer assisted personal interview (CAPI) survey, using a national area probability sample, administered face-to-face by the Census Bureau. The scale of the NCVS makes it difficult for small cities or police jurisdictions to administer, and the cost of a sample that would cover small local areas makes it difficult for the NCVS to produce estimates of victimization at these levels. The ACVS is a self-administered mail survey using an address-based sample (ABS), something that local areas could use.

The ACVS is not a replacement of the NCVS. The NCVS collects detailed information about all victimizations experienced by a respondent over the past six months. A mail survey would be incapable of accommodating the

same scale in detail. The ACVS includes a small subset of the NCVS content. While the NCVS can produce estimates on the incidence and prevalence of various victimization types, the ACVS is well suited for reporting prevalence (or experience) of a smaller set of victimization types. The intended purpose of the ACVS is as a companion to the NCVS for information on victimization collected by local areas.

2.2. Data Collection Design

Our data come from the second year of a large-scale field test. The goal of the year two field test was to assess changes to the mail survey forms, test different data collection approaches, and test the ability of the mail approach to measure change over time. An address-based sample was selected for the field test covering the 40 largest core-based statistical areas (CBSAs) in the U.S. These areas served the goal of testing the mail survey’s potential for producing valid comparison across jurisdictions or local areas. Since victimization is an infrequent event, a large sample was necessary. The total sample for year two of the field test was 217,250 addresses. The protocol for the survey mailings generally followed the mail data collection procedures outlined by Dillman, Smyth, and Christian (2009). This approach included an initial survey mailing, a thank you/reminder post card, and two subsequent nonresponse follow-up mailings. The field test was conducted by Westat with data collection occurring from late September 2016 through early January 2017.

2.3. Experimental Design and Hypothesis

The focus of our paper is on the outcome of experiments conducted in the field test testing the outcome and cost of several combinations of incentives and mail delivery methods. The design crossed three incentive levels as part of the initial survey mailing and two different delivery methods for the last nonresponse follow-up (NRFU). The incentive levels were no incentive (\$0), \$1, and \$2 (provided as cash). The two delivery methods were 1st-class USPS and FedEx. Thus there were 6 incentive-by-mail treatment groups. Table 1 lists the experimental conditions with shortened labels to identify each condition. Sampled addresses were randomly assigned to treatment groups within each of the 40 CBSAs.

Table 1. Incentive by mail delivery method treatment groups

Label	Treatment Condition
0U	\$0 incentive and USPS 1 st class for final NRFU
1U	\$1 incentive and USPS 1 st class for final NRFU
2U	\$2 incentive and USPS 1 st class for final NRFU
0F	\$0 incentive and FedEx for final NRFU
1F	\$1 incentive and FedEx for final NRFU
2F	\$2 incentive and FedEx for final NRFU

We focus on four hypotheses, with two addressing response rates and two addressing cost. We start with the response rate hypotheses:

- H1: Prepaid cash incentives without special delivery methods (Treatments 1U and 2U) will elicit higher overall response than will special delivery methods without an incentive (Treatment 0F).
- H2: Conditions including incentives and special delivery methods (Treatments 1F and 2F) will result in higher response than the conditions without both, or neither (Treatments 0U, 1U, 2U, or 0F), but the combined increase will not be cumulative [e.g. the response rate difference between Treatments 0U and 1U (effect of \$1) plus the difference between Treatments 0U and 0F (effect of FedEx) will be greater than the difference between Treatments 0U and 1F (effect of \$1 and FedEx)].

We base H1 on the idea that an incentive is more motivating than increasing the saliency or importance of the survey through FedEx mail delivery. As suggested by Dillman et al. (2009), prepaid incentives elicit a feeling of reciprocity toward the researcher. A more salient (and costly) delivery method may increase the perceived

importance or the likeliness that a household member sees the survey materials. Additionally, all respondents are exposed to the incentive (assuming opening of the survey package), while only nonrespondents to two prior survey request are exposed to FedEx delivery.

Both incentives and special delivery methods increase mail survey participation. However, we argue that the benefits of each are not cumulative. Our second hypothesis suggests that any increase resulting from FedEx delivery is reduced since many of these respondents were motivated to respond with the prepaid incentive.

Our next two hypotheses focus on cost. Since incentives increase costs of the initial mailing for a given sample size, our metric for comparisons of cost is cost per returned survey. This metric takes into account the cost of fielding the entire sample (for a given condition) divided by the number of returned mail surveys for the condition.

H3: Providing prepaid cash incentives is more cost effective than not using incentives at all (cost per return for Treatments 1U and 2U will be less than for Treatment 0U, and similarly for the F treatments), given a 3-mailings-plus reminder design.

H4: Special mail delivery methods are less cost effective than providing prepaid incentives (cost per return for Treatment 0F is greater than for Treatments 1U and 2U) given a 3-mailings-plus reminder design.

Hypothesis 3 states our belief that including a small prepaid incentive will result in lower cost per return than not providing an incentive. This hypothesized savings is realized through increased response due to the incentive. Through increasing response to the initial mailing, there is a comparable reduction in the level of nonresponse follow-up required.

Special delivery methods for nonresponse follow-up, such as FedEx, can be considerably more expensive overall than initial monetary incentives of \$1 or \$2. When response to early contacts is low, the cost is even greater, since more follow-up is necessary. Hypothesis 4 states that because of these large cost disparities and reduced response without an incentive, relying on special mail delivery methods costs more than providing small prepaid monetary incentives.

3. Results

The effect on response of our incentive and special mailing treatment conditions and relative differences in cost of these conditions is our primary interest. It is equally important to consider what effect our treatment conditions have on the composition of the achieved samples or key outcome measures. We present results in two sections. In the first section, we examine the effects of the experiment on response, achieved sample composition, and key outcomes. Since the focus of the ACVS is measuring victimization, we compare prevalence estimates of different victimization types. We also examine measures on attitudes toward the police. In the second section, we focus on cost per complete for each experimental group. We then look at the change in sample needed without incentives or special mail delivery to achieve the same number of returns, and the total cost of data collection for the new sample.

3.1. Response Rate and Outcome Measures

Table 2 shows the AAPOR RR2 (AAPOR, 2016) response rates for each treatment group. As expected, all conditions that included any treatment (incentive or special mail delivery) resulted in significantly higher response compared to the base condition with no incentive or mailing treatment. The larger incentive, regardless of mailing condition, resulted in higher response. These results confirm findings on the effectiveness of incentives in the survey literature (see Mercer et al. 2015).

Group	Completes	Ineligibles	Nonresponse	Response Rate
0U	9,615	3,200	23,392	29.1
1U	11,249	3,103	21,857	34.0*

2U	12,456	3,040	20,712	37.6*
0F	11,449	2,856	21,903	34.3*
1F	13,077	2,826	20,306	39.2*
2F	13,957	2,753	19,499	41.7*

* $p < 0.001$ when compared with Treatment 0U

Our first hypothesis stated that we expected to observe higher response for our incentive only conditions (without the use of FedEx for nonresponse follow-up) compared to our condition that only used FedEx (no incentive provided). What we observe in Table 2 is that \$1 (Treatment 1U) was about as effective as FedEx delivery alone (Treatment 0F). However, \$2 was more effective than FedEx delivery ($p < 0.001$, *se* 0.37). We find mixed support for our first hypothesis.

For our second hypothesis, we expected a combined increase in response for households exposed to both the incentive and FedEx delivery, but the combined increase would not be cumulative of individual increases. Using our \$1 incentive condition (1U) and FedEx condition (0F) the individual increase for each was 4.9 and 5.2 percentage points, respectively. Our hypothesis states that for the group with both treatments (1F), the observed increase in response should be less than 10.1 percentage points compared to the 0U group. For the \$2 incentive and FedEx delivery (2F), the increase should be less than 13.7 percentage points. We do not observe any strong evidence for our second hypothesis. The \$1 incentive with FedEx delivery is exactly the same as the combined effects with a 10.1 percentage point increase, while the \$2 incentive with FedEx delivery is slightly lower than the combined effects with a 12.6 percentage point increase.

We next reviewed the compositions of the responding sample for each treatment group. We identified five characteristics that correlate with experiencing victimization. Changes in the compositions of these characteristics could lead to changes in key estimates of victimizations. Our characteristics included the following:

- Presence of anyone in the household age 18 – 24
- Presence of a male in the household age 18 – 24
- Presence of anyone of African-American descent in the household
- Presence of anyone of Hispanic ethnicity in the household
- Indication of high mobility – measured through whether reported moved within the last 12 months.

Across all five characteristics we examined, there were no significant differences in achieved sample composition between treatment groups. There were some nominal differences for two characteristics worth mentioning. For presence of anyone African-American, the incentive groups (regardless of delivery methods) showed nominally higher proportions. For presence of anyone Hispanic, the incentive and FedEx combined groups showed nominally higher proportions. While these are only nominal differences, they suggest that treatments to increase response may be bringing in under-represented groups. Our compositional measures are household level, and additional research looking at respondent-level measures may reveal stronger differences.

Another concern is the effect of the incentive and special mail delivery treatments on key outcome measures of victimization. The measures we examine are whether anyone in the household experienced one of three types of victimization. We refer to this exposure as “touched by crime,” measuring victimization prevalence. Our three victimization categories included the following:

- Household reports any serious violent victimization (generally victimizations involving injury)
- Household reports any violent victimization (excludes injury)
- Household reports any property victimization (generally includes only theft of property)

For all three key outcome measures, there were no differences in the proportion of households reporting being touched by (exposed to) these types of victimizations. It is important to note that our victimization measures are prevalence and not incidence estimates. Prevalence is whether a household experienced any incident of a specific victimization type, while incidence measures how many times a victimization type was experienced over the reference period. While significant differences in prevalence were not observed, there may be differences in incidence.

The last measures we review are perceptions of neighborhood safety and attitudes toward police. We examine these measures to determine if the incentive or mail delivery treatments bias responses. For example, attitudes may be influenced by getting money in the mail. Of 13 attitude measures in the survey and we examine 5 that specifically addressed safety and police performance. The 5 questions are listed below; all used 5-point scales.

- On the whole, how much of the time is the community where you live safe?
- Overall, how much of the time is the place where you work safe?
- How would you rate the local police on treating people respectfully?
- How consistent are the local police in applying the laws in the same way to everyone?
- Taking everything into account, how would you rate the job the local police are doing?

We examine the proportion of respondents selecting the top two (most positive) categories for each item. For the first two items, these categories are “always safe” and “mostly safe.” For the last three items, they are “very” and “somewhat” respectful, consistent, or good. Proportions across all treatment groups for each measure were very consistent, showing nearly no variation, with no significant differences. The incentive or FedEx delivery did not have any influence on our attitude measures.

3.2. Cost

Data collection costs are an important factor in determining the value provided by data collection enhancements, such as incentives or more expensive delivery methods. Incentives are a large up-front investment and the return on that investment is not always easy to assess. Special mailings, such as FedEx, can be several times more expensive than 1st class USPS, and are mailed to households that have perhaps indicated refusal to participate by not responding to prior survey requests. There are two approaches to determine cost effectiveness. One is to look at cost per return. This measure takes into account what you get in return by factoring cost across all survey returns. To illustrate this, if we take two equal samples and include an incentive in one, the incentive group will have higher total cost. Due to the positive effect incentives have on response there are more cases that those costs are spread across. The larger the increase in response, the lower cost per return.

Another way to look at this is to determine the change in sample (generally assumed to be an increase) required to achieve the same yield as the treatment condition (incentive, special mail). Using the cost per return of the control group, we can determine how much it would cost to achieve the same yield (i.e., cost for equivalent yield). Essentially, how much would the increase in sample cost to reach the same number of returns as the treatment group? We examine both cost per return and cost for equivalent yield in determining the cost effectiveness of both incentives and FedEx nonresponse follow-up.

For determining data collection cost we included several components including survey printing cost, outgoing USPS 1st class postage, FedEx delivery, incentives, and return survey postage. We excluded survey development cost, which we determined were equivalent across treatments, since the same mail forms were used in each treatment group. Finally, the estimated amount for each component is a hypothetical cost we would expect local areas or jurisdictions to face. Local areas or jurisdictions may not have in-house capabilities such as printing, or be able to receive high volume discounts that larger organizations may enjoy.

Table 3 shows the cost per returned mail survey and the difference in cost per return using the control treatment of no incentive and no special mail treatment (Treatment 0U). Table 3 includes the achieved response rate (AAPOR RR2) and the resulting increase in survey returns for reference. Immediately apparent is the cost savings achieved from both groups with only a prepaid incentive. Including an incentive resulted in a cost per returned survey about 6 percent *less* than that without an incentive, supporting our third hypothesis. When only FedEx delivery was used for final nonresponse follow-up with no initial incentive the cost per return was about 14 percent higher than no special

treatment, showing support for our fourth hypothesis. When incentives were combined with FedEx delivery, the relative cost increase compared to no incentive or special delivery was reduced to about 5 percent more than for Treatment 0U.

Table 3. Response and cost characteristics for each treatment group

Group	Response rate	Increase in Returns (treatment – base 0U)	Estimated Cost/Return	Cost Difference/Return (treatment – base 0U)
0U	29.1	--	\$26.49	--
1U	34.0	1,634	\$24.86	-\$1.63
2U	37.6	2,841	\$24.85	-\$1.64
0F	34.3	1,834	\$30.24	+\$3.75
1F	39.2	3,462	\$27.72	+\$1.23
2F	41.7	4,342	\$27.86	+\$1.37

The results observed in Table 3 show the importance of taking into consideration what you get in return for the additional investment, that is, how many more survey returns. In our experiment, the sample size of each experimental group was just over 36,000. Adding just a \$1 incentive for one group is an immediate added up-front cost of \$36,000 (\$72,000 with a \$2 incentive). The added cost may still be difficult to rationalize, especially when composition of the achieved sample or key estimates are unaffected, suggesting that the gain is more of the same type of respondent. However, this view overlooks two important consequences. The first is the increased efficiency (higher response / getting more with less) from the original sample, which affects the second consequence; increased analytic power (or at the very least meeting analytic power requirements). Another way to look at this is to calculate how much more sample would be required to reach the same yield as our experimental conditions and to then calculate how much more the increase in sample would affect total cost.

We calculated cost for treatment yields using assumptions from our control treatment (no incentive and 1st class USPS mail delivery) and the yield for each treatment group. The control group assumptions were a cost per return of \$26.49, a response rate of 29.1% and an 8% ineligible rate. Our ineligible rate (due to postal nondeliverables) varied across treatments between 7.6 percent and 8.8 percent. The actual rate for our control group was 8.8 percent. The sample size for each experimental group was roughly 36,207.

Table 4. Increased sample requirement for control group to achieve equivalent yield to treatment groups, and resulting comparison of total data collection cost

Comparison	Yield Goal	Sample Needed	Percent Increase	Total Cost Increase (Decrease) vs. Comparison Group
0U to 1U	11,249	41,975	15.9%	\$18,316
0U to 2U	12,456	46,478	28.4%	\$20,439
0U to 0F	11,449	42,721	18.0%	(\$42,913)
0U to 1F	13,077	48,796	34.8%	(\$16,031)
0U to 2F	13,957	52,079	43.8%	(\$19,102)

Table 4 shows how total data collection costs are affected though increasing the sample to achieve similar treatment group yields using our control group assumptions. When compared to conditions that only included a prepaid incentive, total data collection costs increase by over \$18,000 (6.5 percent) versus a \$1 incentive and over \$20,000 (6.6 percent) versus a \$2 incentive. When compared to conditions that used FedEx for final nonresponse follow-up, adding more sample yielded a substantial savings of nearly \$43,000 (12.4 percent) compared to only using FedEx to achieve the same yield. The savings were less (4.4 and 4.9 percent), but still fairly large when compared to treatments that included an incentive and FedEx.

The results in Table 4 support two important observations. The first is that when survey return yield is held constant, including a small prepaid incentive is more cost effective than no incentive or relying on special delivery methods, such as FedEx. The second is that these results should not be misinterpreted to suggest that inefficient uses of a sample (simply adding sample) to reach a targeted number of completes is recommended over interventions to increase response. While we did not observe significant differences in composition of achieved samples or in key estimates, increasing the sample is unlikely to reduce any existing biases any more than incentives or special mail delivery.

4. Discussion

In this paper, we reviewed the results of experiments using both small prepaid monetary incentives and special mail delivery for nonresponse follow-up. Our results align with the large body of research on this topic. Prepaid incentives and FedEx delivery for nonresponse follow-up are both independently effective for improving response and combining both results in larger increases. We examined incentive and mail delivery methods separately, offering insight on their individual and cumulative contributions to response. We also provided a review of the cost effectiveness of each condition rarely seen in the survey literature. In our discussion, we review these results in the context of our four hypotheses and offer some thoughts on the implications of our results.

Our first hypothesis stated that we expected to observe larger increases in response for incentive only conditions when compared to a FedEx only condition. This made sense since the incentive is provided to all sampled addresses, while FedEx is only used for addresses that, due to their nonresponse, are likely to have lower response propensities. Simply put, more people are exposed to the incentive. We observed mixed support for our first hypothesis, finding that prepaying \$1 was equivalent to FedEx nonresponse follow-up. However, prepaying \$2 was more effective than FedEx nonresponse follow-up. This latter finding suggests that incentives are more effective than FedEx nonresponse follow-up alone, but only for amounts in excess of \$1. This is an area in need of additional research.

For our second hypothesis, we believed that while both prepaid incentives and FedEx delivery would have positive effects on response, this effect would not be cumulative when both are used together. We expected that those who may require the use of FedEx to motivate them to respond would have already been motivated by the incentive. We did not find evidence to support this hypothesis. Our experimental conditions that included both an incentive and FedEx resulted in increases in response that were generally equal to the sum of our separate conditions. The implication of this observation suggests that incentives and FedEx delivery may influence different respondent groups. For example, the incentive is likely not to influence respondents who fail to open the survey package, or who feel the importance or legitimacy of the survey is increased with delivery via FedEx. It may thus be important to use both incentives and specialized survey delivery methods. Further research is needed to see if this finding holds for larger incentive levels (e.g., \$5).

We reviewed achieved sample compositions and key outcome measures, finding no differences across any of our experimental conditions. A positive aspect of this result is that our experimental interventions did not result in significant increases in any over-represented groups. However, this also means that our interventions did not significantly increase under-represented groups, such as racial/ethnic minorities, households with low income, or those more likely to be victims of crime. If these interventions are not effective in reaching these groups, then what advantages do they offer? One is the increase in sample efficiency, getting more with less. Another is the potential advantages in reducing data collection cost which we discuss next.

Data collection costs are an important determinant in survey design. The increases in response with prepaid incentives are attractive, but the large up-front cost can be difficult to rationalize and may appear to some to be a waste of (taxpayer) money. In our experiment, we found strong support for our third hypothesis, that small prepaid cash incentives are more cost effective than not providing an incentive. Simply stated, there are more survey returns to spread cost across. While up-front cost may be higher, the cost for each returned survey is lower, given a 3-mailing-plus-reminder protocol. When achieved yields are held equivalent, we see that the large difference in up-front cost disappears.

We also found support for our hypothesis that FedEx delivery for nonresponse follow-up would be less cost effective than offering a prepaid incentive. Again, there is logic to this, since a delivery method, such as FedEx, can

cost five to seven times more than a \$1 incentive per mailing piece. Further, the more nonresponse follow-up needed, the higher the total cost. We observed that combining FedEx with an incentive reduced overall cost compared with FedEx and no incentive, both through increases in the number of surveys to spread cost across and by reducing the number requiring nonresponse follow-up.

Often, researchers feel the need to choose between offering an incentive and using more expensive delivery methods for nonresponse follow-up. Incorporating a different delivery method for nonresponse follow-up is practically a canon of mail data collection, and rightfully so given our results. In the presence of budget constraints, this means incentives are usually the first to go, either rationalized that too many will be discarded in unopened survey packages, are too costly, or the belief in unrealistic response rate assumptions without incentives. We have demonstrated that (at least small) prepaid incentives are an effective means to reduce cost.

References

- The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.
- Biemer, P. P., Murphy, J., Zimmer, S., Berry, C., Deng, G., and Lewis, K. 2018. Using Bonus Monetary Incentives to Encourage Web Response in Mixed-Mode Household Surveys. *Journal of Survey Statistics and Methodology*. 6:2 pp. 240-261. DOI: 10.1093/jssam/smx015
- Church, A. 1993. Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly*. 57:1 pp. 62-79. DOI: 10.1086/269355
- Dillman, D. A., Christenson, J. A., Carpenter, E. H., and Brooks, R. 1974. Increasing Mail Questionnaire Response: A Four-State Comparison. *American Sociological Review*. 39:5 pp. 744-756. DOI: 10.2307/2094318
- Dillman, D. A., Smyth, J. D., & Christian, L. M. 2009. Internet, Mail, and Mixed-Mode Surveys. The Tailored Design Method. 3rd ed. Hoboken, New Jersey, USA: John Wiley & Sons.
- Fay, R. E., Planty, M., and Diallo, M. S. 2013. Small Area Estimates from the National Crime Victimization Survey. Proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA, pp. 1544-1557.
- Hager, M. A., Wilson, S., Pollak, T. H., and Rooner, P. M. 2003. Response Rates for Mail Surveys of Nonprofit Organizations: A Review and Empirical Test. *Nonprofit and Voluntary Sector Quarterly*. 32:2 pp. 252-267. DOI: 10.1177/0899764003251617
- Heberlein, T.A., 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. DOI: 10.2307/2094771
- Mercer, A., Caporaso, A., Cantor, D., and Townsend, R. 2015. How Much Gets you How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly*. 79:1 pp. 105-129. DOI:10.1093/poq/nfu059
- Montaquila, J. M., Brick, J. M., Williams, D., Kim, K., and Han, D. 2013. A Study of Two-Phase Mail Survey Data Collection Methods. *Journal of Survey Statistics and Methodology*. 1. pp. 66-87. DOI: 10.1093/jssam/smt004
- Singer, E., and Ye, C. 2013. The Use and Effects of Incentives in Surveys. *Annals of the American Academy of Political and Social Science*. 645:1 pp. 112-141. DOI: 10.1177/0002716212458082
- Williams, D., Brick, J. M., Edwards, W. S., and Giambo, P. (in press). Questionnaire Design Issues in Mail Surveys of All Adults in a Household. In P. Beatty, D. Collins, L. Kaye, J.L. Padilla, G. Willis, and A. Wilmot (Eds.), *Advances in Questionnaire Design, Development, Evaluation, and Testing*. Hoboken, NJ: Wiley