The Effects of Incentives on the
Consumer Expenditure Interview Survey[*]

Karen L. Goldenberg[1], David McGrath[2], and Lucilla Tan[3]

[1]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., N.E, Washington, DC 20212
[2]Defense Manpower Data Center,1600 Wilson Blvd. Arlington, VA 22209
[3] U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., N.E, Washington, DC 20212

Abstract
Response rates have been falling for the Consumer Expenditure (CE) Quarterly Interview Survey, and CE conducted an experiment to see if an incentive could stem the decline. Since each CE household is interviewed 5 times, another question was whether any effect would persist through the last interview a year later. In this experiment, half the sample received prepaid unconditional incentives (a $20 or $40 debit card) prior to the first interview. Response rates for $40 debit card recipients were 4.5 percentage points higher than those of the control group, and the difference, while smaller, persisted across all 5 interviews. Effects were also noted on some data quality and field contact measures.

Keywords: Incentives; longitudinal survey; panel design

1. Introduction

Response rates to the Consumer Expenditure (CE) Quarterly Interview Survey (Interview Survey) fell from 86 percent in 1990 to 76 percent in 2004 (AAPOR, 2006, response rate 1).[1] Because the survey literature has shown incentives to have a positive effect on response rates, BLS conducted an experiment beginning in November 2005 to determine whether incentives could be used to stem the downward trend. This paper is an abbreviated version of a detailed internal report (Goldenberg et al., 2009). The current document describes the design of the incentives experiment and presents results on response rates, expenditure reporting, and indicators of data quality over five waves of the panel survey. Section 1 introduces the CE Survey program and provides some back-ground from the literature on the use of incentives in surveys. Section 2 describes methodology for the Interview Survey incentives experiment, including the study design, the data, and the analysis methodology. Section 3, "Results," presents the effects of the incentives on response rates, data quality, and sample composition. Section 4 considers the implications of the incentives for field collection costs, and Section 5 summarizes the findings.

---

[*] The opinions in this paper are those of the authors and do not represent official policy of the Bureau of Labor Statistics (BLS). David McGrath was formerly with the CE program at BLS.

[1] AAPOR Response Rate 1 (RR1) is defined as the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus noncontacts plus others) plus all cases of unknown eligibility (unknown if housing unit, plus unknown, other) (AAPOR, 2006).

1.2 The U.S. Consumer Expenditure Survey Program

The U.S. Consumer Expenditure Survey program provides continuous information about how American consumers spend their money. These data are used to support revisions to the Consumer Price Index, to provide annual updates to other Federal agencies for specific purposes, and in many forms of economic research. BLS sponsors the collection of expenditures in two independent surveys. The first is the CE Diary Survey, which collects small, detailed expenditures that respondents record daily. The second survey is the CE Interview Survey, which consists of a series of five interviews designed to obtain detailed information about ongoing and less frequent purchases. Interviewers from the U.S. Census Bureau collect the data for both surveys. Results from the Diary and Interview Surveys are integrated to create published expenditures estimates. The research described in this report is based solely on the Interview Survey.[2]

The Census Bureau conducts about 35,000 interviews across the nation each year for the Interview Survey. Each selected household, or consumer unit (CU),[3] is interviewed five times over a period of 13 months; each of the five interviews is a "wave." The first interview, wave 1, is primarily a bounding interview, designed to limit "telescoping" errors in the wave 2 interview.[4] Data from the first interview do not contribute directly to published expenditure estimates. The second through fifth interviews (survey waves 2 through 5) include questions about expenditures for most non-food purchases, such as housing, furniture, vehicles, insurance, and medical and vacation expenses. In addition, the second (wave 2) and fifth (wave 5) interviews collect data on income and work during the prior 12-month period. The survey was designed to be administered in person, and since 2003 has been conducted by computer assisted personal interview (CAPI). In recent years a substantial proportion of interviews from waves 2 through 5 has been conducted by decentralized telephone interview; this figure averaged 36 percent between 2003 and 2006. One member of the CU usually responds for the entire CU. An interview takes approximately one hour.

The primary measures obtained in the CE Interview Survey are the expenditures reported by respondents. Expenditure questions usually have multiple elements, where respondents are asked to specify the item(s), the quantity, the cost per item, and other details. Research suggests that expenditures are underreported (e.g., Garner et al., 2006; Gieseman, 1987; Silberstein and Scott, 1992; Tucker et al., 2004; 2005). Therefore, CE operates under a premise of 'more is better' in looking at expenditure reports. Associated indicators of data quality in the CE Interview Survey include the number of reported expenditures, the dollar value of those expenditures, and the number of expenditure reports requiring allocation or imputation.

---

[2] BLS conducted a similar experiment in the CE Diary survey between March and November of 2006. See McGrath et al. (2007) for results of the effects of incentives in the CE Diary Survey.

[3] The Interview Survey collects data from consumer units, which include people living in a household related by blood or marriage, or unrelated people who share household expenditures. A household may consist of one or more consumer units. For most housing units, the household and consumer unit are the same. We use the term "consumer unit" and CU in this report.

[4] Telescoping errors refer to reports of purchases from outside the reference period.

1.3 Using Incentives in Surveys[5]

An extensive literature has documented the positive effects of incentives on response rates in surveys.[6] Church (1993) concluded that (1) prepaid incentives are more effective than promised incentives; (2) monetary incentives generate higher response rates than non-monetary gifts; and (3) response rates increase with increasing amounts of money. Although Church focused on mail surveys, a meta-analysis by Singer et al. (1999) determined that incentives are also effective in interviewer-mediated surveys. While the presence of an interviewer should lessen the need for an incentive, Singer et al. reported that the difference between a zero-incentive condition and an incentive condition was statistically significant across a broad range of studies, and that larger incentives resulted in higher response rates. Their data paralleled the Church (1993) findings. At the same time, Singer et al. (1999) note that the higher the initial response rate, the lower the difference between the zero-incentive and incentive condition, and that after controlling for the effects of other variables, the effects of incentives are relatively modest.

Surveys sponsored and funded by the U.S. Government traditionally did not offer respondents incentives as inducements to participate, and government sponsorship of a survey was a major factor in achieving high response rates (Heberlein and Baumgartner, 1978). During the 1990s, however, response rates began to fall in all types of surveys, including those sponsored by the U.S. government. During this period, the U.S. Census Bureau initiated a series of incentive experiments in the Survey of Income and Program Participation (SIPP), a high-burden, face-to-face panel-design interview survey, to see if incentives could help to retain more respondents throughout the length of the survey. The SIPP research demonstrated that incentive effects for large, interview-administered government surveys were similar to those for non-government surveys, and that these effects continued to hold through the 6th interview wave two years after an incentive was provided (James, 1997; Mack et al., 1998).

Most of the incentives studies have been based on cash payments of $1, $2, or $5. However, incentives have gotten larger over time, at least for high-burden surveys. The SIPP studies and later research with the Survey of Program Dynamics (Creighton et al., 2007; Martin et al., 2001) achieved success in retaining respondents and converting former refusals by offering incentives of $20 and $40, and the American Time Use Survey (ATUS) successfully tested incentive amounts of $20 and $40 (Piskurich et al., 2001).

2.   Methodology

The CE Interview Survey incentives experiment was developed and planned by a joint BLS-Census Bureau team. The CE Interview Survey has many similarities with the Census Bureau's SIPP. It involves lengthy, high-burden interviews, was designed to be collected in a face-to-face interview, and is conducted by Census interviewers. In designing the experiment, the team drew heavily on the SIPP experience.

2.1 Study Design

CE conducted the experiment as part of regular production data collection. In this design, half of the wave 1 sample was designated as an incentive condition. CUs in the incentive

---

[5] See Goldenberg et al., 2009, for a more detailed discussion.

[6] Singer (2002) cites numerous review articles on the subject.

condition were sent unconditional prepaid incentives in the form of debit cards with values of $20 or $40; the debit cards were included with the survey's advance letter. The other half of the sample did not receive an incentive and was also split into two groups. CUs were assigned randomly to one of the four treatments. Both the $20 and $40 incentive groups, and one no-incentive group, received advance letters by Priority Mail.[7] The Control Group was not sent an incentive and received its advance letter by standard First Class mail. Incentives were distributed only at the first (wave 1) interview. Regardless of treatment condition, all advance letters in waves 2 through 5 were mailed using standard First Class mail.

All 12 Census Bureau Regional Offices participated in the experiment. About half of each interviewer's wave 1 workload consisted of incentive cases. Table 1 shows planned wave 1 sample sizes prior to any data collection. Because addresses were randomly assigned to the treatment groups, each group has, in expectation, the same proportion of CUs by demographic and geographic variables. The experiment was originally planned to run for a year, but was stopped after 9 months for reasons associated with funding the final debit cards. The actual test included addresses with scheduled wave 1 interviews between November, 2005 and July, 2006. The last incentives were distributed to wave 1 respondents in July, 2006, and CE collected the final wave 5 interview data from those CUs in July, 2007.

Table 1: Experimental Design (Planned Wave 1 Sample Sizes)[*]				
Mail Option for Advance Letter	Incentive Amount			Total
	$ 0	$ 20	$40	
First Class Mail	2,376 (Control)	0	0	2,376
Priority Mail in Wave 1	2,261	2,284	2,282	6,827
Total	4,637	2,284	2,282	9,203

[*] Wave 1 sample addresses, including vacant and demolished units (Census Type B and C housing units), for the planned 12-month collection period. The test was stopped after 9 months.

2.2 Debit Card Assessment Questions
As noted, respondents received the debit cards by mail prior to being contacted for the wave 1 interview. The advance letter mentioned that the cards could be used immediately. BLS added several questions to the CAPI instrument for the incentive version of the wave 1 interview to ascertain receipt and use of the card. If the respondent indicated that the CU had not received a debit card, BLS made arrangements to replace it. As a result of the assessment questions, the interviewers were aware of which CUs received incentives and which did not.

---

[7] Priority Mail ensures additional attention and faster handling by the Postal Service, albeit for a substantially higher fee. Prior research has shown that Priority Mail can be an effective tool in capturing respondent attention to advance materials for a survey, and Census Bureau procedure is to use it for debit card mailings.

2.3 Replacement CUs

The CE Interview Survey samples addresses and not CUs. Some CUs move during the year that their address is in sample. The Interview Survey does not follow CUs after they move, but instead interviews new CUs at the sampled address. The movers-in are called "replacement CUs." If the original CU was assigned to the incentive condition, only the original CU received the incentive, not the replacement CU. For this reason, the analysis excludes the replacement CUs from both the incentive and no-incentive groups.

Table 2 shows the number of eligible occupied housing units in the sample by treatment group and survey wave after excluding replacement CUs. The sample sizes in these tables include all completed interviews, refusals, and noncontacts. Approximately one-fourth of the sample falls into each incentive condition. The 'No Replacement' sample contains about 15 percent fewer CUs, a difference that increases by wave as more CUs move to new addresses. The analyses in this report also exclude addresses that were vacant or that had been condemned or demolished at the time they were visited by the interviewer.

Table 2: Sample Sizes for Response Rate Computations, Excluding Replacement CUs					
Wave	Control Regular Mail	No Incentive Priority Mail	Incentive $20	Incentive $40	Total
1	1,922	1,759	1,838	1,805	7,324
2	1,726	1,599	1,667	1,617	6,609
3	1,610	1,492	1,564	1,521	6,187
4	1,561	1,436	1,512	1,454	5,963
5	1,517	1,395	1,466	1,396	5,774
Total	8,336	7,681	8,047	7,793	31,857

2.4 Measures and Weighting

This paper reports on descriptive statistics by treatment group and interview wave for response rates, expenditures, other data quality indicators, respondent and CU characteristics, and CU income. In most cases data appear separately for wave 1 and as aggregated results for waves 2 through 5. This is because the incentive was distributed in wave 1, and because wave 1 is different from the other four interviews on a number of dimensions. Statistics for waves 2 through 5 represent the average for each experimental group for a variable; i.e., the mean based on the sum across waves 2 through 5 within an experimental group. In some situations wave 1 has been excluded from the analysis, as several of the variables used in this analysis are not available for wave 1.

Weighting for the Interview Survey is performed in stages during the multi-phase CE editing process. Base weights are available prior to editing, non-interview adjusted weights are computed during an early phase, and final calibration weights are computed during the final edits.

- *Unweighted analysis*: Descriptive statistics are based on unweighted data when there is no intent to extrapolate findings to a target population. However, the analysis still

takes into account the Interview Survey's complex design by using PROC SURVEYMEANS in SAS 9.1.[8]

- *Base-weighted analysis:* Response rate calculations use base weights to account for the different probabilities of selection (U.S. Office of Management and Budget [OMB], 2006). The analysis also uses base weights to compute the distribution of categorical demographic variables and income, in order to examine the effect of incentives on the survey's sample composition. SAS Proc SurveyFreq is used for variance estimation to account for the Interview Survey's complex design.

- *Final calibration weighted analysis:* An important component of this analysis is the effect of incentives on expenditure estimates, which are computed for publication using final calibration (replicate) weights. Therefore, the analysis of expenditures by incentive group is based on final calibration weights. Proc Descript in SUDAAN (v9.0, Research Triangle Institute), and the Balanced Repeated Replication (BRR) method of variance estimation, account for the Interview Survey's sample design.[9]

- *Measuring statistical significance:* This report uses two different approaches to test the difference between experimental conditions. A measure between any two of the four treatment groups is considered significantly different at the 5 percent level when the 95 percent confidence intervals of the two groups do not overlap. Demographic variables are compared using the Rao-Scott chi-square, which is a modification of the Pearson chi-square test that takes into account survey design effects.

2.5 Rate Computations

The basis for response, refusal, and noncontact rate computations described in Section 3 is an outcome code assigned to each sample address at each wave of the Interview Survey. The outcome code describes the final disposition of the interview and classifies it into one of four broad categories:

- Completed interview

- Type A - noninterviews (primarily refusals or noncontacts)

- Type B - ineligible, unoccupied, or temporarily vacant sample addresses

- Type C - out of scope units, e.g., buildings that have been condemned or demolished or sample addresses located on a military base.

The response rate is calculated as the number of completed interviews divided by the number of eligible cases (completed interviews plus Type A noninterviews). All response rates shown in this report are comparable to AAPOR RR1 (AAPOR, 2006). The two major components of Type A noninterviews are refusals and noncontacts. The refusal rate is the number of Type A noninterviews due to refusal divided by the number of eligible cases. The noncontact rate is the number of Type A noninterviews due to

---

[8] The CE Interview Survey sample design includes stratification and clustering. We used the PSU as a STRATA variable in SAS PROC SURVEYFREQ. We have determined that this option best approximates the design effect of the Interview Survey's complex sample design.

[9] To reduce the impact of outliers, expenditures below the first percentile of the expenditure distribution for a category were assigned the first percentile value (bottom coded), and expenditures greater than 99 percent of the distribution were assigned the 99 percent value (top coded).

inability to contact divided by the number of eligible interviews. Type B and C cases are ineligible for interview and are not discussed further.

3. Results

3.1 Response Rates

Survey response rates are often used as an indicator of survey quality. High response rates increase the likelihood that the survey respondents represent the target population, which may help to lower potential nonresponse bias. Low response rates decrease the effective sample size, which increases the standard errors used in statistical calculations. In a panel survey such as the CE Interview Survey, high response rates in wave 1 are extremely important because they set the stage for response in subsequent waves.

Response: Table 3 shows that in wave 1 of the panel study—the wave at which the incentives were distributed—response rates increase across the treatment groups. The Control group has the lowest rate, with successively higher response rates for the No Incentive Priority Mail group, the Incentive $20 group, and the Incentive $40 group. The Incentive $40 group is nearly 5 percentage points higher than the Control group ($p<.05$). In addition, the Incentive $40 group response rate is 2.3 percentage points higher than the Incentive $20 group, although this difference is not statistically significant. In later waves, overall response rates decline, but response rates for the Incentive $40 group remain noticeably higher than those for the other conditions and level off at approximately 79 percent. In wave 2, response rates for the Incentive $40 group are only 0.3 percentage points below those of wave 1, and are significantly different from all three of the other groups. There is less variation in the Incentive conditions than in the Control groups in the last two waves of the experiment.

Table 3: Response Rates, by Interview Wave				
Wave	Control Regular Mail	No Incentive Priority Mail	Incentive $20	Incentive $40
1	77.3	78.4	79.5	81.8 [1]
2	76.8	76.1	77.2	81.5 [1,2,3]
3	74.7	74.8	76.0	78.8
4	73.9	73.6	76.5	79.0 [1,3]
5	75.3	75.0	76.4	78.6

Significant difference at p<.05 within a wave: [1] Incentive $40 vs. Control
[2] Incentive $40 vs. Incentive $20 [3] Incentive $40 vs. No Incentive Priority Mail

Since data from waves 2 through 5 contribute to published estimates, it is encouraging to find that the positive effect of incentives on response rates persists from wave 1 to waves 2 through 5. This result is consistent with the research from the SIPP survey by Mack et al. (1998), who found that providing an incentive in wave 1 of a panel survey positively influenced response rates for an extended period of time.

One surprising finding is that most of the response rate differences between the $20 and $40 incentive groups exceed the differences between the Control group and the $20 incentive. The literature shows that providing a small incentive creates most of the effect, and that the rate of improvement diminishes as the incentive increases. In the CE experi-

ment, response rates for the $20 incentive are generally 1 to 2 percentage points higher than those for the Control group in all waves, but the differences are not statistically significant. In short, the $40 incentive performs better than expected, while the $20 incentive does not do as well as anticipated. It is also worth noting that response rates for the No Incentive Priority Mail group are not statistically different from response rates for the Control group in any waves, i.e., it performs about the same as the control group.

While these increases are relatively modest, Singer et al. (1999) noted that surveys with low response rates obtain the largest gains from incentives. The Interview Survey struggles to maintain response rates that are acceptable to OMB, but in the current climate a response rate greater than 70 percent is not "low."

Refusals: "Refusals" are CUs that choose not to participate in the CE survey when the residents are contacted by an interviewer and asked to do so. Participation in the CE Interview Survey normally decreases over the five waves of interviewing (Reyes-Morales, 2003). One question raised in this study is whether incentives have an effect on the number of refusals, as has been shown in other research (Shettle and Mooney, 1999).

Table 4 shows that refusal rates for the Interview Survey increase across waves before levelling off in wave 4 and dropping slightly in wave 5. The incentive does not affect refusals in wave 1; differences in wave 1 refusal rates among the four treatment groups are small and not statistically significant. In waves 2 through 5, however, the magnitude of the difference widens between the groups that did not receive an incentive and the groups that did. In addition, the gap between the $20 and $40 incentives expands. While the differences are generally not statistically significant, the trends are clear: the Control and No Incentive Priority Mail groups have higher refusal rates than either of the Incentive conditions. The Incentive $40 group has a lower refusal rate at every wave, and levels off at about 15 percent after wave 2. These results suggest that providing an incentive in the first wave helped to reduce the refusal rate for the duration of the survey.

Table 4: Refusal Rates, by Interview Wave				
Wave	Control Regular Mail	No Incentive Priority Mail	Incentive $20	Incentive $40
1	13.6	13.5	12.9	12.6
2	16.1	17.3	14.7	13.7
3	18.2	18.8	16.8	15.4
4	19.1	19.3	17.5	15.3 [1]
5	18.9	18.1	18.5	14.9

Significant difference at $p<.05$ within a wave: [1] No Incentive Priority Mail vs. incentive $40

Noncontacts: Noncontacts are occupied CUs whose residents the interviewer is unable to reach for the duration of the field period. Noncontacts lower overall response rates, and to the extent that uncontacted CUs are different from responding CUs, they increase the potential for nonresponse bias. Table 5 shows the noncontact rates for each incentive condition and suggests that the incentive did have an effect on noncontacts. In wave 1, the effect of the $40 debit card is striking, especially when compared to the Control group. The rate for the Incentive $40 group (5.8 percent) is approximately two-thirds that of the Control group's 9.1 percent rate ($p < .05$). The noncontact rates for three of the groups are lower in wave 2 than in wave 1. The Incentive $40 group's rate remains below

that for the Control group (p < .05), and is also lower that that of the Incentive $20 group (p < .05).   Over time, even the $40 incentive loses some of its effectiveness in preventing noncontacts, while the rates for the no-incentive conditions trend downward or remain relatively stable across waves 2 through 5.

		Table 5: Noncontact Rates		
Wave	Control Regular Mail	No Incentive Priority Mail	Incentive $20	Incentive $40
1	9.1	8.0	7.6	5.8 [1,2]
2	7.1	6.6	8.1	4.8 [1,3]
3	7.2	6.4	7.3	5.8
4	6.9	7.1	5.9	5.8
5	5.8	6.8	4.9	6.5

Significant difference at p<.05 within a wave: [1] Control vs. incentive $40
[2]No Incentive Priority Mail versus Incentive $40   [3] Incentive $40 vs. Incentive $20

3.2 Data Quality

Reported expenditures. The primary measure of data quality for the Interview Survey is complete and accurate reporting of the CU's expenditures. Table 6 shows the mean and median expenditures for total expenditures the four treatment groups. These figures represent aggregate reported respondent expenditures for waves 2 through 5. Looking first at mean total quarterly expenditures, respondents in the Incentive $40 group report expenditures about 4.4 percent higher than those in the Control group (difference not statistically significant). Median expenditures do not follow the pattern of increasing across the four treatment groups. An analysis of the 13 detailed expenditure categories that feed into total expenditures shows higher mean expenditure levels among the Incentive $40 group than the Control group in 10 of the 13 categories, but these differ-ences are small and most are not statistically significant (Goldenberg et al., 2009). In short, the incentive had no effect on the level of expenditure reporting.

		Table 6: Median and Mean Quarterly Expenditures		
Expenditures	Control Regular Mail	No Incentive Priority Mail	Incentive $20	Incentive $40
Mean	11,245	11,274	11,653	11,743
Median	8,862	8,486	8,810	9,023

Indirect Data Quality Measures: There are several indirect indicators that provide information about the quality of the interviews. "Good" respondent behaviors include answering more rather than fewer expenditure questions, answering "don't know" or refusing to provide a response less often, consulting records and the CE Information

Book[10] to ensure accurate reporting more often, and answering the income questions with enough information to be labelled a complete income reporter.[11] The cumulative effect of these "good" respondent behaviors is a longer interview and data that requires less adjustment (allocation and imputation) during post-survey processing.

Table 7 shows the results for all of the indirect data quality measures described above where the differences were statistically significant in waves 2 through 5: number of expenditure questions answered, number of questions answered 'don't know' or refused, use of records during the interview, and the effect on processing (available for waves 2 through 5 only). The effect of the incentive treatment on most of the indirect data quality measures is generally small, even when statistical significance is reached. However, overall the incentive groups performed better than the Control group. The other measures examined (mode of interview, use of the CE Information Book during the interview, being a complete income reporter, and length of interview) showed no difference across the experimental groups.

Table 7: Indirect Data Quality Indicators, Wave 1 and Waves 2-5					
Data quality measure	Expected Effect of Incentive	Control group	No Incentive Priority Mail	Incentive $20	Incentive $40
Wave 1					
Expenditure questions answered (number)	More	19.5	19.6	20.1	20.4
Questions answered "Don't know" or Refused (%)	Fewer	5.3	5.2	4.7	4.2[1]
Records consulted during interview (%)	More	29.4	29.0	31.3	29.7
Waves 2 through 5					
Expenditure questions answered (number)	More	45.7	45.0	45.6	47.5[1,2,3]
Questions answered "Don't know" or Refused (%)	Fewer	4.2	4.0	3.7	3.5[1,2]
Records consulted during interview (%)	More	38.8	40.2	41.2	41.8[1]
Expenditures requiring allocation (%)	Fewer	8.5	8.1	8.2	8.0[1]
Expenditures requiring imputation (%)	Fewer	6.3	6.0	5.6[1]	5.5[1]
No data adjustment required (%)	More	83.9	84.7[1]	85.0[1]	85.2[1]

Significant difference p < .05 level: [1] Compared to Control [2] Compared to No Incentive Priority Mail [3] Compared to $20 incentive

---

[10] The Information Book is a spiral-bound notebook containing a series of lists, each of which illustrates different expenditure categories and provides examples of items in those categories, so as to cue respondent recall about purchases.

[11] A complete income reporter is a respondent who provides at least one major source of income (wages or salaries, self-employment income, or Social Security income). However, even complete income reporters may have provide a full accounting of income from all sources, or for all members of the CU.

3.3 Sample Composition and Income

Although the incentive increased the Interview Survey response rates, the effect of the incentives on the composition of the sample is less clear. A series of recent studies demonstrated that nonresponse in CE is not random—in particular, blacks are consistently under-represented, while older people tend to be over-represented (Chopova et al., 2008).[12] However, spending behavior varies by demographic characteristics (McGrath, 2005), so to the extent that incentives help to increase response rates from CUs with characteristics often missed by the Interview Survey, the final sample could reflect the general population more closely than is the case without incentives.

An examination of the demographic composition of the respondents in the treatment groups shows that this did not occur. There were no statistically significant differences across the four experimental groups by gender, race, Hispanic origin, age, or education, either for the wave 1 interview or for waves 2 through 5. One curious (but nonsignificant) pattern is a larger percentage of black respondents in the $20 incentive group than in the control group or the $40 incentive group. Similarly, there were no statistically significant differences for CU size, for whether the CU owned or rented the residence, or residence in or out of a metropolitan statistical area. There was a nonsignificant pattern of more respondents from single-person CUs, and more metropolitan statistical area residents.

Expenditures are generally correlated with income, but as shown in Table 6, expenditures did not vary by incentive treatment. Neither did income. Table 8 shows that the means, both pre- and post-imputation, increase across the treatment groups (not statistically significant), while the medians are roughly the same for both the control group and the $40 incentive group. In addition, quintiles of income are roughly similar across the treatment groups.

Table 8: Mean and Median Income by Treatment Group[*]				
Characteristic	Control Regular Mail	No Incentive Priority Mail	Incentive $20	Incentive $40
Complete income reporter in waves 2 and 5 (%)	86.1	84.4	86.4	87.0
Complete Income Reporters For Waves 2 and 5:				
Pre-Imputation Income: Mean before tax	$59,596	$60,764	$61,121	$61,239
Median before tax	$45,000	$43,800	$44,800	$45,025
All responding CUs for Waves 2 and 5				
Post Imputation Income Mean before tax	$64,189	$65,029	$66,110	$67,117
Median before tax	$49,672	$47,000	$48,885	$49,368

[*] Comparisons of pre-imputation income are based on complete income reporters only. During data editing, components of income are imputed, so all responding CUs are included in the post-imputation income comparisons. As a result, post-imputation figures are based on a larger number of CUs. Differences across treatment groups are not statistically significant,

---

[12] The study showed that nonresponse bias is not a significant concern for reported expenditures in the CE Interview Survey, even though there are demographic differences in response patterns.

4. Field Collection Costs

One argument for providing incentives is that the incentive might influence respondents to cooperate with the survey more readily, thus reducing the number of contacts needed to complete a case, and consequently field costs (Singer 1999). If incentive cases require fewer contacts to complete, this would lead to a reduction in field costs that could at least partially offset the cost of the incentives.

To some extent, the data support this expectation. Table 9 shows several measures related to the average number of contact attempts per case, as recorded in the Contact History Instrument. These measures appear separately for wave 1 and averaged across waves 2 through 5.

Table 9: Contact Attempts by Incentive Group, Wave 1 and Waves 2-5								
Contact Attempt Characteristics	Wave 1 (Mean)				Waves 2-5 (Mean)			
	Control group	No Incentive Priority Mail	Incentive $20	Incentive $40	Control group	No Incentive Priority Mail	Incentive $20	Incentive $40
Total Number of Contact Attempts	4.3	4.3	4.3	4.1	4.0	3.9	3.9	3.8[1]
Number of attempts by personal visit	3.3	3.4	3.3	3.1[2]	1.8	1.8	1.8	1.7[1,2]
Total number of contacts	1.9	1.9	1.9	1.9	1.8	1.8	1.8	1.8
Days between 1st and last attempt	11.1	11.0	10.9	10.2[1]	10.3	10.2	10.0	9.8[1]

Significant difference at $p<0.05$: [1]Compared to control [2]Compared to No Incentive Priority Mail

The Incentive $40 group had the smallest average number of contact attempts and attempts by personal visit in wave 1, although there is no difference in the number of actual contacts. The same effects hold for waves 2 through 5. For both statistics, the Incentive $40 group required significantly fewer attempts than the Control group ($p<.05$) in waves 2 through 5. Personal visit attempts are important because they are an important factor in field costs, as each attempt "costs" interviewer time and mileage to reach the sample address, but may or may not result in an interview. Incentives also affected the length of the field period. In wave 1, it took about a day less to complete $40 incentive cases than control cases, an average of 10.2 days between the first contact attempt and the final disposition of the case for the Incentive $40 group compared to 11.1 days for the control group ($p < .05$). The difference was about a half day in waves 2 through 5 (an average of 9.8 days for the Incentive $40 case compared with 10.3 days for a control case, $p<.05$).

Unfortunately, CE does not have detailed information on Census field data collection costs for the period of the incentives test, nor any field cost data which allow direct comparisons of incentive and no-incentive cases.

5. Discussion

The most important results from the CE Interview Survey experiment are that response rates were higher in the incentive treatment conditions, that the Incentive $40 group had higher response rates and fewer noncontacts than the Incentive $40 group, and that providing respondents with incentives only in wave 1 of this panel survey appears to have created an environment that stayed in effect throughout all 5 waves of the survey. By the final interview, it had been a year since respondents received an incentive, but compared to respondents in the Control groups, the incentive recipients exhibited positive response behaviors on a number of dimensions.

In addition to response, incentive recipients performed better when compared to the Control group on most of the indirect data quality measures: they answered more expenditure questions, they used records more frequently, they provided fewer don't know and refused answers, and their reported data required fewer imputations and allocations. The $40 incentive performed better than the Incentive $20 group on 7 of 11 measures, but some of the differences were very small and most were not statistically significant.

Incentive respondents reported slightly higher levels of expenditures overall and for most spending categories, although differences were modest and generally did not attain statistical significance. They were also more likely to be complete income reporters, although there are no statistically significant differences across the experimental treatment groups in reported income or on demographic characteristics.

The total number of contact attempts needed to resolve a case, and the number of contact attempts by personal visit, are both components of overall data collection costs. Respondents receiving the $40 incentive required fewer overall contacts, and fewer personal visit attempts, than the Control group, and less field time to resolve their cases.

One limitation to this analysis is the absence of appropriate data with which to evaluate the cost of implementing incentives for the CE program. An understanding of the cost implications of incentives is important, as the literature suggests that incentives could "pay for themselves" once the initial investment is in place. But apart from this limitation, the CE program believes the incentive experiment was successful, and that the survey would benefit from the introduction of a $40 incentive into the regular data collection program.

References

American Association for Public Opinion Research (AAPOR). 2006. *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys.* Lenexa, KS: AAPOR.

Chopova, Boriana., Jennifer Edgar, Jeffrey Gonzalez, Susan King, David McGrath, and Lucilla Tan. 2008. Assessing nonresponse bias in the CE Interview Survey: A summary of four studies. Washington, DC: U.S. Bureau of Labor Statistics, internal report, August.

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

Creighton, Kathleen P., Karen E. King, and Elizabeth A. Martin. 2007. "The Use of Monetary Incentives in Census Bureau Longitudinal Surveys." Research Report Series, Survey Methodology, #2007-2, January. Washington, DC: U.S. Census Bureau.

Garner, Thesia, George Janini, William Passero, Laura Paszkiewics, and Mark Vendemia. 2006. "The CE and the PCE: A Comparison." *Monthly Labor Review* 129 (9), September.

Gieseman, Raymond. 1987. "The Consumer Expenditure Survey: Quality control by comparative analysis." *Monthly Labor Review*. March.

Goldenberg, Karen L.; David McGrath, and Lucilla Tan. 2009. "The Effects of Incentives on the Consumer Expenditure Interview Survey. Washington, DC: U.S. Bureau of Labor Statistics, internal report.

Heberlein, Thomas and Robert Baumgartner 1978. "Factors affecting response rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review* 43:447-462.

James, Tracy. 1997. "Results of the wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation." *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

Mack, Stephen, Vicki Huggins, Donald Keathley, and Mahdi Sundukchi. 1998. "Do Monetary Incentives Improve Response Rates in the Survey of Income and Program Participation?" *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association, 529-34.

Martin, Elizabeth, Denise Abreu, and Franklin Winters. 2001. "Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation." *Journal of Official Statistics* 17: 267-284.

McGrath, David E. 2005. "Comparison of Data Obtained by Telephone versus Personal Visit Response in the U.S. Consumer Expenditures Survey." *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

McGrath, David E. 2006. "An Incentives Experiment in the U.S. Consumer Expenditure Quarterly Survey." *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

McGrath, David E., McDermott, N., Chopova, B., and Gomes, C. 2007. "The Effects of Incentives On the Consumer Expenditure Diary Survey." Washington, DC: Bureau of Labor Statistics, internal report.

Piskurich, Karen, Dawn Nelson and Diane Herz. 2001. "Maximizing Respondent Contact in the American Time Use Survey." Presented at the American Association of Public Opinion Research (AAPOR) Conference. Montreal, Canada.

Reyes-Morales, Sally E. 2003. "Characteristics of Complete and Intermittent Respondents in the Consumer Expenditure Quarterly Interview Survey." *Consumer Expenditure Survey Anthology 2003*. Washington, DC: Bureau of Labor Statistics:25-29.

Shettle, Carolyn and Geraldine Mooney. 1999) "Monetary Incentives in Government Surveys." *Journal of Official Statistics* 15:231-50.

Silberstein, Adriana R. and Stuart Scott. 1992. "Seasonal Effects in the Reporting of Consumer Expenditures." *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

Singer, Eleanor, Nancy Gebler, Trivellore Raghunathan, John Van Hoewyk, and Katherine McGonagle. 1999. "The Effect of Incentives in Interviewer-Mediated Surveys." *Journal of Official Statistics* 15 :217-30.

Singer, Eleanor. 2002. "The Use of Incentives to Reduce Nonresponse in Household Surveys." In Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, eds., *Survey Nonresponse*. New York: John Wiley & Sons, Inc.

Tucker, Clyde, Paul Biemer, Brian Meekins, and Jennifer Shields. 2004. "Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A Micro-Level Latent Class Analysis." *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

Tucker, Clyde, Paul Biemer, and Brian Meekins. 2005. "Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A further Micro-Level Latent Class Analysis."

JSM Proceedings, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

U.S. Office of Management and Budget (OMB). 2006. *Standards and Guidelines for Statistical Surveys.* Washington, DC.