Journal of Official Statistics, Vol. 17, No. 2, 2001, pp. 267–284

Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation

Elizabeth Martin[1], Denise Abreu[2], and Franklin Winters[1]

Panel attrition due to nonresponse is a serious problem for longitudinal surveys because it reduces sample representativeness and may bias estimates. This article reports the results of an incentive experiment that targeted prepaid monetary incentives to nonresponding households from a prior round of interviewing. Households were randomly assigned to receive a debit card worth 20 USD or 40 USD, or no incentive. 20 USD and 40 USD both significantly improved conversion rates of prior noninterviews. Households in the high poverty stratum were more responsive to 20 USD than households in the low poverty stratum; race and marital status also interacted with incentive effects. Interviewers' notes for experimental cases were coded and analyzed to examine motivational influences on respondents' reactions to incentives. Results show that burden concerns expressed in a prior interview were associated with an announced intention to stop participating in the survey, which led to higher attrition subsequently. Incentive effects were no different for respondents who had complained about the survey's burden than those who had not.

Key words: Nonresponse bias; nonresponse conversion; burden; interviewer notes; methods experiment; longitudinal survey.

1. Introduction

Panel attrition due to nonresponse is a significant problem for longitudinal surveys, such as the Census Bureau's Survey of Income and Program Participation (SIPP), because it reduces the representativeness of survey estimates, and may bias them. Households in poverty have higher attrition rates than nonpoverty households in the SIPP (Waite, Huggins, and Mack 1997).

The SIPP introduces a new panel of sample households every three years. Interviewers return to sample households every four months to collect detailed income and employment information. Panels typically run for two and a half years, but the panel introduced in 1996 was extended to four years, to improve the reliability of estimates (Guarino et al. 1999). To maintain response rates and reduce attrition bias, the U.S. Census Bureau

[1] U.S. Census Bureau, Washington, DC 20233, U.S.A. E-mail: emartin@census.gov
[2] Formerly U.S. Census Bureau, currently National Agricultural Statistics Service, U.S. Department of Agriculture.
Acknowledgements: This article reports the results of research and analysis undertaken by U.S. Census Bureau staff. It has undergone a U.S. Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. We thank Melinda Crowley, Eleaner Gerber, Ashley Landreth, and Gloria Prout, who, with the first author, coded the interviewers' notes. A copy of the coding scheme is available on request. We thank John Bushery, Larry Cahoon, Jeffrey Moore, Bill Nicholls, and Karen King for helpful comments.

planned and implemented an experimental incentives program. Despite the bureau's best efforts, the 1996 SIPP panel had even higher attrition rates than usual, as measured by the permanent sample loss rate, the proportion of households dropped from the survey due to nonresponse in the first wave, or two successive noninterviews in subsequent waves. After seven rounds (or waves) of interviewing, permanent sample loss had reached 29.9 percent, higher than in the case of any previous panel. The U.S. Census Bureau attempted to reduce attrition by implementing an additional incentives experiment which directly targeted nonrespondents.

Nonresponding households were randomly assigned to receive a prepaid incentive of 20 USD, 40 USD, or no incentive in the subsequent round of interviewing. In this article, we analyze the effects of the incentives upon conversion rates, and examine the characteristics of households which were most responsive to incentives. We also use information provided by interviewers' notes recorded in a prior interview to explore motivational factors which may influence effects of incentives. We address the following questions:

- Do incentives improve conversion rates?
- Does the amount of an incentive play an important role in increasing conversion rates?
- Are incentives equally effective for refusals and other noninterviews?
- Do low income households react better to incentives? What other demographic characteristics predict conversion rates, and responsiveness to incentives?
- Do respondents' motivations and concerns influence their response to an incentive?

2. Related Research

In a recent literature review, Singer et al. (1999) conclude that the well-documented positive effects of incentives in mail surveys also hold in surveys conducted by interviewers in person or by telephone. The positive (but modest) effects of incentives appear to hold for fresh respondents, panel respondents, and nonrespondents. Cash is more effective than a gift, even holding constant the value of the gift, and prepaid incentives are more effective than promised or contingent incentives. However, a promised incentive is better than none.

Consistent with the literature, an incentive experiment conducted at the initial contact for the 1996 SIPP panel concluded that incentives significantly increased response rates. Primary sample units were randomly assigned to receive no incentive, or 10 USD or 20 USD in the form of paper vouchers. Incentives were introduced as a "token of appreciation" by SIPP interviewers at the initial contact. Householders were told to expect a check in the mail two to three weeks after they filled in their name and returned the voucher to the U.S. Census Bureau in a postage-paid envelope. Receipt of an incentive was not conditional on giving an interview. For the most part, interviewers were assigned to only one treatment group and were aware of the experiment. James (1997) found that, compared to 10 USD or nothing, the 20 USD incentive significantly reduced initial nonresponse. Overall, Wave 1 nonresponse was 7.5 percent for the 20 USD group compared to 9.1 percent for both the control and 10 USD groups; the reduction occurred in both high and low poverty areas. The average number of hours required to complete a case was lower in the incentive group compared to the control group. Mack et al. (1998) found that the Wave 1 incentive effect held up in subsequent waves of interviewing. Compared to the control group, the 20 USD group had a lower rate of total sample attrition for Waves 2–6. The reduction occurred for both poverty and nonpoverty households. 20 USD reduced household nonresponse, person nonresponse, and item-level nonresponse for the amount of gross pay. A ''booster'' incentive of 20 USD given in Wave 7 to all low income households that had received a Wave 1 incentive also appeared to contribute to a reduced nonresponse rate (Sundukchi 1998). (See Creighton, King, and Martin, 2001, for a summary of incentives use in the U.S. Census Bureau's longitudinal surveys.)

A high rate of attrition may occur because the SIPP collects detailed income information that is often difficult to provide, requiring considerable time and effort from respondents. Thus, Singer et al.'s (1999) finding that survey burden does not significantly interact with the effectiveness of incentives is particularly germane to the SIPP. As Singer and her colleagues note, this finding is somewhat contrary to expectation, since on theoretical grounds one might expect incentives to be more effective (even necessary) when a survey imposes a large burden upon respondents. In this article, we examine the influence of respondents' concerns about survey burden (as expressed to and recorded by interviewers) upon response to an incentive.

Social exchange theory provides the most common theoretical framework for interpreting the effects of incentives on survey participation (see, e.g., Dillman 1978; Groves, Cialdini, and Couper 1992). The norm of reciprocity (Gouldner 1960) obliges one who receives a benefit to repay it at some time, and thus explains why receiving an incentive leads to greater cooperation with a subsequent request to participate in a survey. Considerations of social exchange also suggest that individuals will provide help in proportion to the gain that is expected or has been received, in order to maintain equity in the relationship.

However, the relationship between incentives and survey cooperation is probably more complex than simple notions of quid-pro-quo might suggest. Incentives may have unintended effects on intrinsic motives for survey participation, such as altruism, civic duty, or interest. Research shows that civic-mindedness is highly predictive of decennial census participation (see, e.g., Couper, Singer, and Kulka 1998). If a monetary incentive leads respondents to attribute their survey participation to the reward, then future intrinsic motivation may be reduced (Bem 1965) and continued cooperation may depend on continued payment of rewards. In the survey context, Singer et al. (1997) find that students who cooperated with a survey request after receiving a small gift perceived themselves as motivated by interest, while those given 10 USD attributed their participation to the incentive.

Singer, Van Hoewyk, and Maher (1998) note that payment of an incentive may create an expectation of future payments. Such expectations may reduce future cooperation in a panel survey if they are not met. Singer, Van Hoewyk, and Maher found experimentally that respondents who received an incentive were more likely than those who did not to agree that ''people should get paid for doing surveys like this.'' However, they were also more, not less, likely to cooperate with a subsequent survey request with no offer of an incentive. The authors note that respondents may have interpreted the earlier payment as covering their subsequent participation as well. It is possible, and consistent with social exchange theory, that payment of an incentive leads respondents to calculate what contribution on their part is proportional to the incentive they have received. We examine this hypothesis below.

A second, methodological purpose of this article is to explore interviewer notes as a source of information about the interview situation and respondent motivations. We build on Groves and Couper's conclusion that ''utterances ... by householders are informative about the likelihood of eventual cooperation'' (1998: p. 265). Based on analysis of systematic observations recorded by interviewers, they found that ''negative and time delay statements from householders in one contact portend lower cooperation in later contacts and ... at the final disposition of the sample case'' (1998: p. 165). They argue for the development of reliable, cheap indicators of householder behavior that might be used in response propensity models. Interviewers' notes are a readily available, cheap source of information about householders' utterances which do not require a special data collection effort. However, interviewers' notes are less systematic than observation data of the sort collected by Groves and Couper, and must be coded to be usable. Here, we code the notes and explore their use as predictors of cooperation in a conversion attempt. If measures derived from them are predictive, then interviewer notes may represent a source of data that could be fruitfully exploited for research and operations purposes.

3. Facts about the SIPP

The SIPP's main objective is to provide policy makers and others with accurate and comprehensive information about levels and determinants of income and program participation of persons and households in the United States. The SIPP data are used to help formulate and evaluate initiatives in welfare reform, tax reform, and the improvement of entitlement programs such as Social Security and Medicare.

The survey covers the noninstitutionalized population of the United States, including people living in housing units as well as group quarters, such as dormitories, rooming houses, and family-type housing on military bases. People living in military barracks and institutions, such as prisons and nursing homes, are excluded.

The SIPP is a nationally representative longitudinal survey with a multi-stage sample design. Its design consists of two strata within each Primary Sampling Unit (PSU) – one for households above and another for households below 150 percent of the poverty threshold. Strata were formed using 1990 census data for housing units and blocks (Siegel and Mack 1995). Households in the high poverty stratum are oversampled within each PSU.

The 1996 Panel was introduced in April 1996 and continued through March 2000, by which time each sample household should have been interviewed every four months for four years, for a total of 12 times. A total of 36,730 households was interviewed in Wave 1. To spread out interviewing and processing workloads, each round (or wave) of interviews is divided into four subsamples or rotation groups. Interviews for each wave take place over four months, with one rotation group interviewed each month (U.S. Census Bureau 1999). All household members 15 years old and over are interviewed by self-response, if possible; proxy response is permitted when household members are not available for interviewing. Interviews take 30 minutes per person, on average.

The current practice in the SIPP is to revisit nonrespondents one more wave after their initial nonresponse. Wave 1 nonrespondents are an exception because they are not followed up in subsequent waves. On average, about a third of household nonrespondents in one wave are converted to interviews in the next. Noninterviews in two consecutive waves result in a household being dropped permanently from the survey.

4. Method and Design of Experiment

The experiment was conducted in Waves 8 and 9. The sample consisted of about 2,800 households that were nonrespondents in Waves 7 or 8, but had been interviewed the previous wave. Included were noninterviews due to refusal, no one home, temporary absence, language problem, or for reasons other than sample ineligibility. Four sample selection strata were created by cross-classifying poverty stratum (high versus low) by noninterview category (refusal versus other noninterview). After defining stratum boundaries and sorting units by geographical region, three random subsamples of almost equal size were selected with each assigned to one of three treatment conditions – a 20 USD incentive, a 40 USD incentive, or no monetary incentive (control group). Assignment to treatment groups was independent of the Wave 1 treatment assignments. Note that the design does not match or control for the demographic characteristics of households assigned to the four experimental groups. In Section 5.2, we control for demographic characteristics and examine their effects.

Consistent with current nonresponse conversion procedures, all groups received an advance letter prior to the interviewer's visit. The letter received by the control group was the usual conversion letter sent to nonrespondents, while the letter received by the incentive groups also included information about the incentive, and a debit card and PIN number. Letters to all groups were sent by priority mail, to ensure that householders received the incentives. Priority mail is a departure from the usual procedure of sending letters to nonrespondents by first-class mail. Debit cards were used to reduce interviewer procedural error, and because research shows that prepaid incentives are more effective than promised ones. Interviewers were not blind to incentive treatment, and provided an incentive at the door to householders who claimed they did not receive the letter with the prepayment.

The regular SIPP instrument was administered by Computer Assisted Personal Interviewing in face-to-face interviews, with telephone follow-ups conducted to obtain missing information.

An additional source of data is provided by interviewers' notes. A sample of cases was used to develop a coding scheme to capture information about the interview situation. For each of 33 types of events or statements, ''1'' was recorded if the event/statement was mentioned in the note, and left blank for no mention; thus, any number (or none) of the codes might apply to a given note. Interviewer notes were coded by five coders, with an eight percent random sample double-coded to evaluate reliability and adjudicate code interpretations. The level of coder disagreement was about 2.5 percent. Coders were blind to interview outcomes and incentive treatment.

Due to the limited availability of staff for this exploratory component of the study, only one wave's worth of notes (half the experimental cases) were coded. Interviewer notes were coded for 1,284 complete or partial interviews in Wave 6 which became noninterviews in Wave 7 and were included in the incentive experiment in Wave 8. 75 percent of these cases were assigned one or more codes. An additional 293 cases in the Wave 8 incentive experiment were not coded; most were outmovers from sample households for which Wave 6 notes were unavailable.

It is an open question whether interviewer notes provide meaningful data that can be analytically useful. Notes are recorded by interviewers for their own use and to communicate information to the regional office or other interviewers who may be assigned a case. Their content is highly variable, and may include contact information, descriptions of problematic situations or changes in a household, ''venting,'' requests or advice about the handling of a case, problems with the automated instrument, etc. The notes are not rigorous, systematic observations, and the absence of a comment does not mean an event did not occur.

5. Analysis and Results

Section 5.1 summarizes the basic results of the incentive experiment. (Abreu and Winters (1999) report slightly different results obtained using a less edited version of the data.) Section 5.2 examines demographic predictors of conversion rates and the responsiveness of different groups to incentives. Section 5.3 uses coded information from interviewers' notes to consider motivational influences on conversion rates and incentive effects.

CPLX (Fay 1988) is used to fit log-linear models to cross-classifications of various independent variables and the dependent outcome variable. Weighted estimates are computed using SIPP base weights which are the inverse of the probability of selection. Alternative hierarchical models are compared to select a model that is best-fitting in the sense that it cannot be significantly improved by adding additional effect parameters, nor can effects be dropped without a significant loss of fit. To arrive at a best-fitting model, a forward model selection procedure similar to that described by Goodman (1971) is used: a model including all two-way interactions involving the dependent variable was found to have acceptable fit and used as the baseline model, and the significance of three-way (or higher order) interaction terms involving the dependent variable is assessed by comparing the goodness-of-fit of the baseline model with the same model modified to include the interaction in question. All models are constrained to fit the joint distribution of the independent variables (see Goodman 1971). Jackknifed variance estimates and jackknifed chi-squared test statistics are computed to take into account the complex design of the survey. Jackknifed Pearson chi-squared test statistics (X^2) are used to evaluate goodness of fit of alternative models, and jackknifed Likelihood-ratio chi-squared test statistics (L^2) are used to compare models and evaluate the contribution of particular terms to the model (see Fay 1985; 1988, p. 10.1). The α level used is .10, the U.S. Census Bureau's standard.

It must be kept in mind that the experiment is restricted to households interviewed in Wave 6 or Wave 7 that became noninterviews the next wave and were included in the incentives experiment the subsequent wave. These sample restrictions limit the generalizability of the results.

Table 1. Conversion rates for incentive groups

	Group	Conversion rate (percent)
Wave 7		41.0
Waves 8/9	Control (priority mail only)	45.9
	20 USD incentive	51.2
	40 USD incentive	54.1
	Total	50.5

5.1. The effects of experimental design variables

Table 1 presents Waves 8 and 9 conversion rates, calculated as the number of interviews divided by the number of interviews plus noninterviews, for the experimental cases. (Noninterviews include refusals, not-at-homes, language barrier, temporary absences, and first time Type D's, i.e., households that moved to an unknown address or outside a SIPP PSU. Demolished, condemned, and vacant units, those under construction, and two-time Type D's are excluded, as are about ten control group cases which were given an incentive after learning of the study.) The experimental cases were at risk of attrition, since those not converted in Wave 8/9 were permanently dropped from the survey. That is, the attrition rate for Wave 8/9 is the inverse of the conversion rate in Table 1.

Differences among incentive treatment groups are statistically significant ($X^2 = 2.66$, d.f. $= 2$, $p < .004$). Both 20 USD and 40 USD obtained significantly higher conversion rates than the control group, while the 20 USD and 40 USD groups do not differ significantly. Also shown is the conversion rate in Wave 7, before the experiment began. The control group's conversion rate is significantly higher than the Wave 7 rate (45.9 versus 41.0 percent), suggesting that priority mail alone improved conversion. This inference is uncertain, because we do not know what the rate for control group would have been without priority mail. (The rate for the control group also is higher than the conversion rates of 30.8, 35.1, 29.7, and 38.5 percent for Waves 3–6, respectively.)

To further examine the effects of the design variables on conversion rates, we fit log-linear hierarchical models to the five-way cross-classification of the design variables (Noninterview type, Poverty stratum, and Wave), experimental treatment variable (Incentive), and dependent Outcome variable (1 = interview, 2 = noninterview). The model is constrained to fit the joint distribution of the independent variables. Table 2 presents β coefficients for the best-fitting model; coefficients more than twice their standard errors are shown in bold.

For dichotomous variables, the single effect shown is the difference between the effect of the first category and the average effect. For variables with more than two categories, the parameter shown is the difference between the effect of that category of the variable and the average effect (see Fay 1988). Positive values indicate that a category (or combination of categories) had positive effects on conversion rates.

The main effect parameters indicate that;

– the control group had significantly lower conversion rates than either incentive group,

Table 2. Logistic regression coefficients for four predictors of conversion

Predictors of conversion		β	S.E.
Incentive amount	0 USD	**−0.226**	0.059
	20 USD	0.124	0.067
	40 USD	0.102	0.069
Noninterview type [1 = Refusal in prior wave, 2 = other noninterview]		**−0.409**	0.046
Poverty stratum [1 = high, 2 = low]		0.075	0.041
Poverty X Noninterview type*		**0.096**	0.044
Poverty X Incentive†	0 USD	−0.067	0.055
	20 USD	**0.157**	0.067
	40 USD	−0.090	0.072
Goodness of fit: Jackknifed Pearson $X^2 = -.69$, d.f. $= 16$, $p > .5$			

†$p < .10$, *$p < .05$, **$p < .001$.

- refusals from a prior wave had a lower conversion rate than other noninterviews, and
- conversion rates for the two poverty strata did not differ significantly.

There is no Wave x Outcome term, because conversion rates did not vary between Waves 8 and 9, nor did wave interact with other variables in affecting outcome. The model includes two significant interaction terms. Incentives and poverty interacted in their effects on conversion rates, as did poverty and refusal status. Because the model is hierarchical, inclusion of these interaction terms implies that lower order terms are also included (for the same reason, the separate contribution of the lower order terms to the model cannot be evaluated). The model fits the data very well, with $X^2 = -.69$ on 16 d.f. and $p > .50$. Results are discussed below.

Conversion rates by incentives and poverty

Households in the high poverty stratum were differentially responsive to 20 USD (as indicated by $\beta = .157$ for the high poverty/20 USD group in Table 2). As Table 3 shows, conversion rates for the control group were similar in the low and high poverty strata. Incentives improved conversion rates in both strata, but not uniformly.

In the high poverty stratum, a higher conversion rate (61.1 percent) was achieved with 20 USD than with 40 USD (54.9 percent), but the difference is not significant. Both rates are significantly higher than the 47.1 percent rate for the control group ($p < .001$ and .081, respectively). In the low poverty stratum, a higher rate was achieved by 40 USD than by the control or 20 USD ($p < .007$ and .064, respectively), while 20 USD did not significantly improve conversions.

Table 3. Conversion rates (in percent) for incentive groups by poverty stratum

Poverty stratum	Incentive treatment			
	Control group	20 USD	40 USD	Total
High poverty stratum	47.1	61.1	54.9	54.6
Low poverty stratum	45.7	48.8	54.0	49.5
Total	45.9	51.2	54.1	50.5

In other words, 20 USD (or more) improved the conversion rate in the high poverty stratum, but 40 USD was needed to boost the rate in the low poverty stratum. (As discussed below, this result may be due to the indirect effects of marital status and race, which correlate with poverty.)

The interaction implies that the composition of interviewed households varied among incentive groups: 23.2 percent of households interviewed in the 20 USD group were in the high poverty stratum, compared with 18.8 and 19.5 percent of households in the control and 40 USD groups, respectively; the difference is significant ($p < .056$). However, the inference that relatively more poor households were interviewed in the 20 USD group is uncertain because stratum correlates imperfectly with poverty at the household level. (Mack et al. (1998) report Wave 1 household poverty rates of 27 and 11 percent in high and low poverty strata, respectively.)

Conversion rates by type of noninterview

Conversion rates are much lower for prior refusals than for other noninterviews, as reflected by the significant negative coefficient ($\beta = -.409$) in the model. The results in Table 4 seem to suggest that incentives were more effective in converting refusals than other noninterviews.

Compared to the control, 40 USD increased the conversion of refusals by 10.7 percentage points, but only 4.1 percentage points for other noninterviews. This difference is not reliable (i.e., the interaction involving the three variables is insignificant, $p < .39$). Thus, statistically, the effect of incentives did not differ for refusals and other noninterviews.

Finally, the best-fitting model also includes an interaction between poverty, noninterview type, and outcome ($\beta = .096$), indicating that conversions were higher among high poverty refusals, relative to the average effect. The rate of conversion of prior refusals was 48.3 percent in the high poverty stratum and 39.8 percent in the low poverty stratum (these results not shown). The rate of conversion of other noninterviews was similar in both strata (about 64 percent).

In sum, analysis of the effects of the experimental design variables shows:

- Incentives improved conversion rates, with no overall difference in the effect of 20 USD versus 40 USD.
- The effects of incentives within the two poverty strata were not uniform;
  - 20 USD yielded improvement in the high poverty stratum, with no further gain from 40 USD, and
  - only 40 USD had an effect in the low poverty stratum.
- Results are inconclusive on the differential efficacy of incentives for prior wave

Table 4. Conversion rates (in percent) for incentive groups by type of noninterview in prior wave

Noninterview type	Wave 7	Incentive treatment			
		Control	20 USD	40 USD	Total
Refusal	29.8	36.1	41.1	46.8	41.4
Other noninterview	58.1	61.1	66.1	65.2	64.2
Total	41.0	45.9	51.2	54.1	50.5

refusals and for other noninterviews; only the effect for refusals is statistically significant.
- Priority mail alone appeared to improve conversion rates compared to the usual procedure.

5.2. Demographic predictors of conversion

In this section, demographic variables are introduced to examine whether incentives were more effective in some groups than others. The main effects of householder's race, education, marital status, poverty stratum, noninterview type, and incentive on conversion rates are estimated by fitting log-linear models to the cross-classification of these seven variables. We also test for interactions among demographic variables, incentive treatment, and outcome. Missing demographic information was obtained from a prior interview, when available. Rates of missing data are 7.5 and 18.7 percent for race and marital status, respectively. (Marital status is dichotomized; ''single'' includes widowed, divorced, separated, and never married persons.)

Table 5 presents logistic regression coefficients for the best-fitting model predicting conversion. The model is constrained to fit the joint distribution of the independent variables Race, Education, Poverty, Marital Status, and Noninterview type. Incentive

Table 5. Logistic regression coefficients for demographic predictors of conversion

Predictors of conversion			β	S.E.
Incentive amount		0 USD	−0.178	0.101
		20 USD	0.165	0.111
		40 USD	0.013	0.101
Noninterview type [1 = Refusal, 2 = other noninterview]			**−0.423**	0.050
Race		White	−0.053	0.093
		Black	0.018	0.105
		Other	0.035	0.142
Education [Some college versus none]*			**−0.117**	0.053
Marital status [1 = Single, 2 = married]			−0.089	0.048
Poverty stratum [1 = high, 2 = low]			0.043	0.056
Poverty X Noninterview type*			**0.098**	0.048
Incentive amount X Marital status*		0 USD	−0.037	0.061
		20 USD	**0.157**	0.059
		40 USD	**−0.121**	0.059
Incentive amount X Race*	0 USD	White	0.030	0.099
		Black	−0.091	0.131
		Other	0.061	0.172
	20 USD	White	**−0.238**	0.107
		Black	−0.182	0.137
		Other	**0.420**	0.186
	40 USD	White	0.209	0.112
		Black	0.273	0.142
		Other	**−0.482**	0.182
Goodness of fit: Jackknifed Pearson $X^2 = .70$, d.f. = 216, $p = .23$				

†$p < .10$, *$p < .05$, **$p < .001$.

Amount is allowed to vary with respect to the other independent variables. The model provides an acceptable fit to the data ($X^2 = .70$, d.f. $= 216$, $p = .23$).

In addition to the effect of noninterview type (discussed above), education has a negative effect on conversion rates: householders with some college or more were less likely to be converted to interviews than those who had never attended ($\beta = -.117$).

The model also includes two interaction effects, in addition to the interaction between Poverty, Noninterniew Type, and Outcome (discussed above). The interaction between Incentive Amount, Marital Status, and Outcome occurs because conversion rates for single householders are significantly enhanced in response to 20 USD ($\beta = .157$) and depressed in response to 40 USD ($\beta = -.121$), relative to the effect for married householders. In other words, 20 USD was sufficient to improve response for single individuals, but it took 40 USD for married ones.

Another interaction ($p < .057$) involves Race, Incentive Amount, and Outcome. Although conversion rates do not vary by race, 20 USD was less effective for whites ($\beta = -.238$) and more effective for individuals of ''Other race'' ($\beta = .420$) than it was on average. 40 USD was less effective for ''Other race'' individuals ($\beta = -.482$) than on average.

After controlling for the interactions involving race and marital status, the interaction between poverty stratum, incentive treatment, and outcome (shown in Table 2) is not significant ($p < .18$) and drops out of the model. The differential responsiveness of the two poverty strata to 20 USD may have been due to the indirect effect of marital status and race. When they are controlled, there is no statistical evidence that households in the two strata responded differently to incentives. Alternatively, it is possible that household-level poverty influences responsiveness to incentives, but the stratum-level measure is too crude to detect the effect. Race and marital status may be picking up the effects of household poverty, with which they are correlated.

In summary:

- College-educated householders were less likely to be converted to interviews.
- Marital status and race did not directly influence conversion rates, but did interact with incentive effects;
  - single householders were more responsive to 20 USD than married ones, and
  - white householders were less responsive to 20 USD, and ''other race'' householders were more responsive to 20 USD and less responsive to 40 USD, compared to the average effect.
- Household-level poverty is correlated with race and marital status, and may well account for the interaction effects; its role remains uncertain.
- Differential responsiveness to incentives resulted in compositional effects: a larger fraction of the households interviewed in the 20 USD group was in the high poverty stratum.

5.3. Motivational influences on response to incentives

Of all the reasons for dropping out of a survey, concerns about burden seem those most directly addressed by an incentive. This is suggested explicitly by some respondents who made clear that the survey requires more time and effort than should be expected gratis. For example, one respondent told an interviewer ‘‘she feels the government takes advantage of persons by not giving some compensation for time spent,’’ and another is quoted as saying, ‘‘This is such an imposition on us for only the 10 dollars you gave us the first time. I think we should be paid every time we do this. I have had enough of your questions.’’ If burden concerns are addressed by compensation, one might expect incentives to be more effective among those who had complained about burden prior to the offer of an incentive. However, an incentive may backfire if its amount is deemed inadequate for the amount of time and effort required by a survey, as suggested by the second quote above. An offer of money may invite a calculation of whether the incentive amount provides adequate compensation for time and effort spent and lead to a more explicit link between perceived burden and the decision to participate.

We are interested in several questions. Do measures derived from interviewers' notes predict the outcome of later interview attempts? Were people who complained about survey burden in Wave 6 more responsive to incentives in Wave 8? Is there evidence of long term effects of the Wave 1 incentive on householder motivation? For instance, did receipt of an incentive in Wave 1 lead to a heightened concern about survey burden?

We introduce two indicators derived from interviewers' notes. BURDEN indicates any complaint about survey burden, including length of interview, duration of survey, repetitiousness of questions, dislike of questions, general mentions of being tired of SIPP, and/or complaints of being taken advantage of or receiving no personal benefit. QUIT indicates a comment that the respondent wanted or intended to quit the survey. About ten percent of Wave 6 notes for the experimental cases recorded a complaint about burden, and the same fraction recorded a desire to quit. Wave 1 incentive treatment is included in the analysis. (Wave 1 incentives were administered only in Rotations 2–4, so Rotation 1 cases are added to the control group that received no incentive.) Table 6 presents results of a loglinear analysis of interrelations among five variables: Wave 1

Table 6. Logistic regression coefficients for three response variables

Effects		β	S.E.
Outcome × Wave 8 Incentive	0 USD	−0.090	0.073
	20 USD	−0.082	0.075
	40 USD	**0.171**	0.085
Outcome × QUIT**		**−0.439**	0.101
BURDEN × QUIT		**0.344**	0.064
QUIT × Wave 1 incentive	0 USD	−0.093	0.094
	10 USD	−0.051	0.105
	20 USD	0.143	0.095
BURDEN × Wave 1 incentive	0 USD	−0.024	0.082
	10 USD	−0.031	0.096
	20 USD	0.055	0.079
BURDEN × QUIT × Wave 1 incentive*	0 USD	−0.059	0.085
	10 USD	−0.164	0.086
	20 USD	**0.223**	0.080
Goodness of fit: Jackknifed Pearson $X^2 = -0.41$, d.f. = 50, $p > .5$			

†$p < .10$, *$p < .05$, **$p < .001$.

Incentive (0 USD, 10 USD, 20 USD); Wave 8 Incentive (0 USD, 20 USD, 40 USD); BURDEN (1 = Wave 6 interviewer note records a complaint about survey burden; 2 = no mention); QUIT (1 = Note records respondent's desire or intent to stop participating in the SIPP, 2 = no mention); and Outcome (1 = Interview in Wave 8, 2 = Noninterview). BURDEN, QUIT, and Outcome are treated as dependent (or response) variables, while Wave 1 and 8 incentives are independent variables.

The model includes the Wave 8 incentive × outcome effect (discussed above), although with only half the sample the overall effect does not quite reach significance ($p < .108$); only the coefficient for the 40 USD treatment is statistically significant. There is no evidence that the incentive was more effective for householders who had previously complained about burden than for those who had not. (That is, there is no significant interaction between BURDEN, outcome, and Wave 8 incentive.) Nor is there evidence that BURDEN directly affected conversion rates.

Stating an intention to quit is highly predictive of subsequent attrition: 70.5 percent of respondents who in Wave 6 declared they would quit were true to their word and ended up as noninterviews in Wave 8, compared to 49.6 percent of respondents who did not mention quitting. That is, the odds on attrition more than doubled among respondents who said they were quitting. The fact that respondents carried out their intentions with fair reliability means that their statements provide prognostic information that may usefully inform conversion strategies.

The predictive power of QUIT also holds when Wave 7 outcome (refusal versus other noninterview) is introduced into the model (these results are not shown). Our results thus support Groves and Couper's (1998) conclusion that the types of statements made by householders are informative about the likelihood of eventual cooperation, beyond the information contained in call-level result codes. However, more general samples are needed to examine the generalizability of these results. We do not know whether similar statements made by Wave 6 respondents who were interviewed in Wave 7 also predict Wave 8 outcomes.

The Wave 1 incentive did not directly influence Wave 8 outcome, nor is there evidence that recipients of an incentive in Wave 1 later were more likely to complain about burden or say they were quitting. However, the Wave 1 incentive did affect the relationship between BURDEN and QUIT, as indicated by a significant three-way interaction term in the model.

Table 7. Percent intending to QUIT, by BURDEN complaint and Wave 1 incentive treatment

Wave 1 incentive treatment	Percent who say they intend to quit among burden complainers and non-complainers		Unweighted N
	Burden complainers	Non-complainers	
0 USD	21.6	8.2	658
10 USD	19.5	10.8	302
20 USD	43.7	7.3	324
Total	26.2	8.5	1,284

BURDEN and QUIT are highly correlated ($\beta = .344$): 26.2 percent of those respondents who complained about survey burden, also said they wanted to quit the survey, compared to 8.5 percent of those who did not complain about burden. Moreover, the association between these two variables is intensified by receipt of 20 USD in Wave 1, as shown in Table 7.

Around 20 percent of those in the control and 10 USD groups who complained about survey burden said they intended to quit SIPP. This fraction rose to almost 44 percent in the 20 USD group, representing an increase of more than three-fold in the odds on deciding to quit.

In other words, respondents who complained about survey burden were more than three times as likely to declare their intention to quit the survey **if** they had previously received 20 USD. And, as noted above, declaring an intention to quit more than doubled the odds on attrition.

This finding is consistent with the hypothesis that a monetary incentive leads respondents to explicitly calculate the time and effort that is proportional to the amount they have received, and to link that calculation to the decision to continue (or stop) participating. If so, our results suggest that 20 USD, but not 10 USD, led respondents to more clearly link perceived burden with a decision to quit. (Perhaps 10 USD is too small to invoke an expectation of compensation.) More research is needed to clarify the link between incentive amount, respondent expectations, and the decision to participate.

An alternative explanation is that the Wave 1 20 USD incentive was effective in retaining respondents who were reluctant to participate in Wave 1, so the 20 USD group includes more reluctant respondents at Wave 6. This interpretation is consistent with evidence that attrition was lower for the 20 USD group up through Wave 6 (see Mack et al. 1998). However, note that in Table 6 there are no differences in complaints about burden among the Wave 1 incentive groups; rather, it is the association between QUIT and BURDEN that is influenced by the Wave 1 incentive.

The results of the modeling suggest the following causal chain: perceived burden is associated with a decision to stop, which leads to higher attrition in Wave 8. Thus, perceived burden did not directly influence Wave 8 attrition, but rather had an indirect effect through an influence upon the decision to participate. Because the association between perceived burden and intention to quit is intensified by receipt of 20 USD in Wave 1, the results also imply that the Wave 1 incentive indirectly influenced Wave 8 attrition (although the effect is not significant).

In sum:

- Respondents who declared in Wave 6 that they were quitting the survey were more than twice as likely to attrit in Wave 8.
- A complaint about burden was positively associated with a declared intention to quit, and the association was stronger if respondents had received 20 USD in Wave 1.
- Expressed concern about burden did not directly influence conversion rate, nor did it influence response to incentives. Wave 8 incentives were equally effective among householders who complained about burden and those who did not.

6. Conclusions

The results of the 1996 SIPP Panel Waves 8 and 9 Incentive Study reveal that offering incentives to prior wave nonrespondents substantially improved conversion rates. A larger incentive amount (40 USD versus 20 USD) did not yield significantly higher conversion rate overall. Use of priority mail to send a conversion letter also appears to have improved conversion rate.

Conclusions about overall effects of incentives are modified by analyses showing that incentives interact with householders' demographic characteristics. The results are inconclusive due to the lack of a measure of household level poverty in these data, which may account for many of the interaction patterns found. Several population groups (householders who were single, or other race, or in the high poverty stratum) were responsive to a 20 USD incentive, while others (householders who were married, or white, or in the low poverty stratum) were only responsive to a 40 USD incentive. Intuitively, it makes sense that poverty should influence incentive effects, since a smaller monetary incentive should be worth more to a poor household that is lacking in resources. However, the inference that a household's poverty level influences its response to an incentive is uncertain, and understanding the underlying causes of these differential incentive effects requires additional analysis using a household level poverty measure.

If their effects vary across demographic groups (as they appear to), incentives may influence the demographic makeup of the interviewed sample. In this study, the demographic composition of interviewed households varied between the 20 USD and the other conditions, even though the overall effect of the 20 USD incentive was no different from the 40 USD incentive. Such interactions may complicate any decision about the use of incentives. Our results underscore the need to analyze incentive effects for subgroups as well as the total sample, and to carefully identify possible differences in sample composition which may arise from their use. The survey designer must be sensitive not only to the overall effects of incentives, but also to their potential effects on reducing or increasing sample bias. In the case of the SIPP, previous research has documented that attrition is greater for poverty than nonpoverty households. Thus, an incentive which is differentially effective in poverty households might help correct the bias resulting from attrition by retaining more poverty households in sample. This reasoning, and consideration of results such as those in Table 3, might lead a survey designer to prefer a 20 USD incentive, even if a higher incentive amount led to a larger overall improvement in the attrition rate. If (as was the case in this study) incentives do not affect households uniformly, then their effects upon sample composition and survey results may be complex. Such complexities can only be disentangled by evaluating incentives using experimental designs.

A second question our results do not clearly answer is whether incentives are equally effective for all types of nonresponse, or only for refusals. Additional data are needed to provide sufficient numbers to estimate the incentive effects for different noninterview types. The answer to this question could help researchers optimize the design of incentive programs by targeting them to groups for which they are most effective.

The results provide mixed findings concerning the hypothesis that householders' motivations may determine their response to incentives and may in turn be affected by them. We reasoned that concerns about burden are directly addressed by an offer of an incentive. However, the effects of Wave 8 incentives were not conditioned by respondents' expressed concerns about burden, or their declared intentions to stop participating in the survey. This result is consistent with Singer et al.'s (1999) conclusion that incentives do not interact with burden in their effects on response rates. While the Wave 8 incentive did not interact with perceived burden in affecting conversion rates, the earlier Wave 1 incentive did have an effect. The prior 20 USD incentive apparently strengthened the resolve to quit the survey among respondents who complained about burden, although the effect (if any) upon attrition appears slight. Possibly, this result occurred because a 20 USD incentive in the first interview introduced the notion of a quid pro quo, and evoked an explicit link between calculation of the level of effort appropriate for the incentive and the decision to continue participating. It is also probable that the Wave 1 20 USD incentive retained more reluctant respondents, who were then somewhat less likely to be converted after a noninterview many waves later.

Our preliminary analysis suggests that the motivational effects of incentives may not be homogeneous. The effects on initial and cumulative response rates are positive, as shown by analyses of the SIPP Wave 1 and Wave 8 incentives experiments. At the same time, incentives also may have complex effects on subsequent motivation. Our results provide a caution that the motivational effects of incentives may play out over time, and may not be immediately apparent through examination of simple associations between incentive treatments and initial response rates. Additional experimental research is needed to examine short and long term effects of incentives on respondents' motivations in longitudinal surveys such as the SIPP.

The results of coding and analyzing interviewers' notes suggest that this material represents a rich (and readily available) source of information about the interview situation that might be more effectively used both for research and operations purposes. An indicator based on Wave 6 notes was highly prognostic of outcomes several waves later, although this result needs to be replicated on more general samples before being accepted as certain. This finding supports Groves and Couper's (1998) conclusion that householders' statements and behavior to interviewers predict their subsequent cooperation. Currently, information from interviewers' notes is used in an ad hoc way by interviewers and supervisors. In fact, research shows that interviewers are more effective in their conversion attempts when they have access to descriptive call records from previous interview attempts (Ahmed and Kalsbeek 1998). However, more systematic use of narrative material found in interviewers' notes may allow field staffs to improve their conversion rates and develop customized conversion strategies that more effectively address particular nonresponse motives and situations than is possible on an ad hoc basis. More effective use of this material (and, perhaps, more training for interviewers on what to record in their notes) may yield new insights about situational and motivational influences on survey cooperation.

7. References

Abreu, D. and Winters, F. (1999). Using Monetary Incentives to Reduce Attrition in the

Survey of Income and Program Participation. Proceedings of the American Statistical Association, Survey Research Methods Section.

Ahmed, W. and Kalsbeek, W. (1998). An Analysis of Telephone Call History Data from the Behavioral Risk Factor Surveillance System. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas Texas, August.

Bem, D. (1965). An Experimental Analysis of Self-Persuasion. Journal of Experimental Social Psychology, 1, 199–218.

Couper, M., Singer, E., and Kulka, R. (1998). Nonresponse in the 1990 Census: Politics, Privacy, and/or Pressures. American Journal of Politics, 26, 59–80.

Creighton, K.P., King, K.E., and Martin, E.A. (2001). The Use of Monetary Incentives in U.S. Census Bureau Longitudinal Surveys. Pp. 289–310 in Statistical Policy Working Paper 32: Seminar on Integrating Federal Statistical Information and Processes. Federal Committee on Statistical Methodology, Office of Management and Budget. Washington, DC.

Dillman, D.A. (1978). Mail and Telephone Surreys: The Total Design Method. New York: Wiley.

Fay, R.E. (1985). A Jackknifed Chi-Squared Test for Complex Samples. Journal of the American Statistical Association, 80, 389, 148–157.

Fay, R.E. (1988). CPLX Program Documentation. Washington DC: U.S. Bureau of the Census.

Goodman, L.A. (1971). The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications. Technometrics, 13, 1, 33–62.

Gouldner, A. (1960). The Norm of Reciprocity: A Preliminary Statement. American Sociological Review, 25, 2, 161–178.

Groves, R., Cialdini, R., and Couper, M. (1992). Understanding the Decision to Participate in a Survey. Public Opinion Quarterly, 56, 475–495.

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

Guarino, J.A., Huggins, V.J., Fay III, R.E., and Dajani, A.N. (1998). Variances for the Survey of Income and Program Participation (SIPP) 1984–1996 Panels. Proceedings of the American Statistical Association, Survey Research Methods Section.

James, T. (1997). Results of the Wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation. Proceedings of the American Statistical Association (Survey Research Methods Section).

Mack, S., Huggins, V., Keathley, D., and Sundukchi, M. (1998). Do Monetary Incentives Improve Response Rates in the Survey of Income and Program Participation? Proceedings of the American Statistical Association, Survey Research Methods Section, 529–534.

Siegel, P.H. and Mack, S.P. (1995). Overview of Redesign Methodology for the Survey of Income and Program Participation. Proceedings of the American Statistical Association, Survey Research Methods Section.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., and McGonagle, K. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. Journal of Official Statistics, 15, 217–230.

Singer, E., Gebler, N., Van Hoewyk, J., and Brown, J. (1997). Does $10 Equal $10? The Effect of Framing on the Impact of Incentives. Paper presented at the annual meeting of the American Association for Public Opinion Research, Norfolk, VA.

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

Sundukchi, M. (1998). SIPP 96: Wave 7 Incentives. U.S. Census Bureau Memorandum from Baer to Kirkendall, April 1.

U.S. Census Bureau (1999). SIPP Quality Profile – May 1999. SIPP Working Paper #230. Washington DC: U.S. Census Bureau.

Waite, P.J., Huggins, V.J., and Mack, S.P. (1997). Assessment of Efforts to Reduce Nonresponse Bias: 1996 Survey of Income and Program Participation (SIPP). Paper prepared for presentation at the 8th International Workshop on Household Survey Nonresponse, Mannheim, Germany, September.

Received February 2000
Revised December 2000