Journal of Official Statistics, Vol. 27, No. 2, 2011, pp. 279–299

Effects of Increasing the Incentive Size in a Longitudinal Study

Willard L. Rodgers[1]

To assess the possible impacts of increasing the incentive size on response rate and data quality, at one wave of a longitudinal study sample members were randomly assigned to receive one of three incentives: the same $20 incentive per person they had received in previous waves; $30; or $50. Those offered $50 had a higher response rate than those offered $20, and this positive impact persisted for at least the next four waves of biennial data collections. There is some evidence that higher incentives improve data quality as indicated by probabilities of failing to answer questions and of giving responses that would reduce the interview length.

Key words: Panel attrition; data quality; respondent incentives; nonrandom missing data.

1. Introduction

Research on the response rates to mail surveys has shown that respondent incentives are one of two design characteristics, along with the number of mailings, that have consistently and substantially increased the response rate (Heberlein and Baumgartner 1978; Yu and Cooper 1983). A meta-analysis of published research findings with respect to the effects of incentives on nonresponse rates in telephone and face-to-face interview surveys (Singer et al. 1999b) demonstrates that incentives increase the response rates to these types of surveys as well.

Most of the research on incentive effects, regardless of mode, has looked at response rates in cross-sectional surveys, or studies with one or two follow-up waves. For panel studies, especially those intended to continue for many waves, achieving high response rates in the follow-up waves is at least as important as achieving a high initial response rate, since the cumulative effect of attrition across multiple waves can be devastating. Moreover, the value of each respondent in a panel study cumulates over time, making investigation of methods to reduce attrition a sensible methodological component of such studies; and, should higher incentives prove to be effective in reducing attrition, the cost of those increased incentives could prove to be a useful investment in the long-term viability of such studies. A review published in 2002 (Singer 2002) indicated that little research had been done on the usefulness of incentives for maintaining high response rates in panel

[1] Survey Research Center, University of Michigan, Ann Arbor, MI 48106-1248, U.S.A. Email: wrodgers@umich.edu
Acknowledgments: Major funding for the Health and Retirement Study is provided by the National Institute on Aging. Many useful suggestions were made by four reviewers and the editor.

studies, or on the optimal size of such incentives. Considerably more has been done in the years since that review, as summarized by Laurie and Lynn (2009). The general finding has been that introducing an incentive for the first time after several waves with no incentive, or increasing the size of the incentive, improves the response rate. For example, at Wave 14 of the British Household Panel Survey (BHPS), a random subsample was given an increased incentive (£10, after having received £7 for eight waves, for adults; £5, an increase from £4, for youths). Among sample members who were interviewed at the previous wave, the response rate was considerably higher (96% vs. 93%) for those given the higher incentive; and for reluctant respondents, defined by having failed to give an interview at the previous wave, the response rate was over twice as high (13% vs. 6%) for those given the higher incentive. A recent paper (Zagorsky and Rhoton 2008) looked at the effects of offering an incentive to sample members of the National Longitudinal Study of Women (NLSW) for the first time after some 20 waves of data collection in which no incentives were offered. The primary finding was that incentives of either $20 or $40 were effective in increasing the response rate among sample members who were nonrespondents at the previous wave.

There is evidence that unconditional incentives, given to the sample members before they have agreed to be interviewed, are more effective than incentives given only after the completion of the interview. Singer et al. (1999b) reviewed data from 39 experimental studies of incentive effects in telephone and face-to-face surveys, and report that in studies that offered a fixed incentive size but varied whether they were given conditionally or unconditionally, those in the unconditional groups consistently had higher response rates than those in the conditional groups. Castiglioni et al. (2008), however, failed to confirm this pattern in a three-wave panel study in which one group was at each wave given an unconditional 10€ voucher, a second group was promised a 10€ voucher if they agreed to be interviewed, and a third group was offered no incentive.

There is also evidence that incentive effects endure beyond the wave at which they are proffered. For example, Mack et al. (1998) report on an experiment incorporated into the 1996 panel of the Survey of Income and Program Participation (SIPP), offering evidence that an incentive offered at the initial wave of a longitudinal study has positive effects on response rates, not only at the baseline wave, but also at each of five subsequent waves.

Another possible impact of incentives is with respect to the quality of the data provided by respondents. Arguments can be made for expecting both positive and negative consequences. On the one hand, larger incentives could engender a stronger sense of obligation to take the interview seriously and therefore expend greater effort to answer all of the questions accurately. Dillman (1991; 2000) has developed the Total Design Method (and its successor, the Tailored Design Method) for mail surveys based on social exchange theory, on the basis of which sample members are predicted to be more likely both to participate and to answer questions thoughtfully if the perceived benefits outweigh the perceived costs. On the other hand, higher incentives could persuade individuals with low intrinsic motivation to participate, and such individuals may tend to provide poorer quality data than other respondents. Hansen (1980) found that sample members in a mail survey who were offered small incentives were more likely to participate than were those offered no incentive, but that the data provided by those given the incentives were of poorer quality. He interprets this pattern in terms of self-perception theory: those who participate

without an external incentive are more likely to perceive themselves as good survey participants, and therefore take the task more seriously.

The empirical evidence is mixed. While Jäckle and Lynn (2008) reported that respondents who were randomly assigned to receive an incentive at Wave 2 of a longitudinal study of youth had higher item missing data rates than those who received no incentive, neither Ryu et al. (2005) nor Petrolia and Bhattacharjee (2009) found evidence of a reduction in data quality among those given incentives. Singer et al. (1999b) reported that across thirteen studies that examined the quality of the data, using as criteria either item missing data rates or the average number of words recorded in response to open-ended items, none showed a deleterious effect of incentives on data quality and six showed positive effects. Finally, in the study most directly relevant to the one reported in this article, Zagorsky and Rhoton (2008) found that among sample members in the National Longitudinal Study of Mature Women who refused to participate in one wave of interviews but who were interviewed at the following wave, those who were offered an incentive at the second wave had longer interviews, and were asked more questions, than those who were not offered an incentive. They interpret this finding as an indication that the incentives improved the quality of the data, noting that since these were people who had already been through many waves of data collection, they had learned that affirmative responses to many questions (for example about short-term jobs or relationships) resulted in their being asked additional questions.

Telephone and face-to-face surveys are two-way interactions between interviewers and respondents, raising the possibility that effects of incentives on response rates and data quality may be mediated by the interviewers. Knowledge of incentives may change interviewers' behavior, perhaps making them more confident that sample members who have received (higher) incentives will be cooperative (see Singer 2002; Castiglione et al. 2008). Singer et al. (2000) reported on an experiment designed to distinguish interviewer-mediated effects from other effects of incentives; they found no evidence of a difference in response rates between two groups, one for which the interviewers were blinded, the other for which the interviewers were informed that sample members had received a $5 check along with an advance letter. (Both groups had considerably higher response rates than a third group that received no incentive.)

An obvious impediment to providing incentives to sample members is the cost of doing so. These costs must be weighed against the costs of other survey methods (e.g., the number of call backs or offering alternative modes for participation) that have been found to be effective in increasing the response rate and/or improving the quality of the data. Moreover, Singer et al. (2000) found that in a telephone survey fewer calls were needed to finalize sample members who were offered a $5 incentive than those given no incentive, suggesting that incentives may actually reduce overall costs of data collection. James (1997) reported that sample members in the 1996 SIPP panel that were given $20 incentives required fewer interviewer visits than those given no incentive. Rodgers (2002) found that the direct costs of higher incentives in the 2000 Wave of the Health and Retirement Study (HRS) were at least partly offset by reduced interviewer hours and expenses, especially for those who were not interviewed at the previous wave.

This article describes an experiment conducted as part of one wave of the HRS, which is a longitudinal survey of a nationally representative sample of persons age 51 and older,

including their spouse or partner regardless of age. Interviews have been conducted biennially since 1992. The general practice for data collections conducted from 1993 through 1998 was to include a check for $20 along with an advance letter that was sent to each sample member (including spouses and partners) prior to their being contacted, by telephone or face-to-face, by an interviewer. The experiment, done as part of the 2000 data collection, consisted of changing the amount of the incentive check for randomly selected sample members, with the objective of examining the effect of incentive size on the response rate, on the amount of effort required to finalize the case, and on the quality of the data obtained from respondents. In this article, we examine the effects of the size of the incentive on the response rate on the immediate wave and on subsequent waves, and also examine the effects on item missing data rates.

2. Methods

2.1. Design of the Health and Retirement Study (HRS)

HRS began in 1992 as a longitudinal survey of a sample of individuals born between 1931 and 1941, selected using an area probability sample design. The starting point for the design was a sample of 69,337 housing units. Of those, 9,419 units (13.6%) were found to be nonsample (either unoccupied or non-households). All but 214 (0.36%) of the identified housing units were screened for eligibility: whether or not any household member was born between 1931 and 1941, inclusive. The eligibility rate was 15.5%. A household member was selected in each eligible housing unit, using a random selection procedure if more than one member was born between 1931 and 1941. If the selected person was married, or living in a marriage-like situation, the spouse or partner was also asked to participate, regardless of year of birth, and with respect to data collection procedures was treated identically to the initially selected person. In about a third of the selected housing units (3,037), an unmarried (and unpartnered) individual was selected, whereas in the remaining two-thirds (6,230) of the housing units a couple was selected, so that the total number of individual sample members was 15,497.

In 1998 the original HRS panel was merged with that of a parallel study, Asset and Health Dynamics among the Oldest Old (AHEAD), which in 1993 started following a sample of those born before 1924; and these two samples were supplemented with samples from the birth cohorts of 1924–1930 (the "Children of Depression Age," or CODA sample) and those of 1942–47 (the "War Babies," or WB sample). The AHEAD sample was based on a dual frame, with most of it drawn from the same household screening that formed the basis for the HRS sample, but with half of those born in 1913 or before drawn from the Medicare Enrollment Database (MED; for details, see Rodgers 1996). The CODA sample was drawn entirely from the MED, and the WB sample was drawn entirely from the 1992 household screen. (In 2004 and 2010, samples of the birth cohorts of 1948–1953 and those of 1954–1959, respectively, were interviewed for the first time, but since these samples were added after the incentive experiment was conducted in 2000 they will not be considered in this article.) Details on the design of the HRS and AHEAD studies and on the baseline response rates are provided elsewhere (Juster and Suzman

1995; Soldo et al. 1997; updates are available at the HRS website: http://hrsonline.isr.umich.edu/).

Reinterview data collections have been conducted every two years through 2010. Sample members were asked to participate at each follow-up wave if either they or their spouse or partner participated at the baseline interview. That is, response at a follow-up wave is not a condition for eligibility at future waves. Exceptions are made only when a sample member or his or her spouse or partner asks to be permanently removed from the study, and these are kept in the denominator when response rates are calculated.

On average, about 6% of the sample members alive at one wave have died by the following wave two years later. Final interviews are sought with proxies (generally the surviving spouse or an adult child) following the death of a sample member. In addition, some of the interviews for living sample members have been conducted with proxy informants, often because the sample person has physical or cognitive limitations that made their participation difficult or impossible, but also if a sample person was unavailable or refused to do an interview but was willing to let someone else answer the questions.

The respondent burden has been considerable, whether measured in terms of the frequency with which they have been asked to participate in a data collection, the average length of the interviews, or the difficulty of the questions they are asked to answer. The median length of nonbaseline interviews conducted in both the 1998 and 2000 waves was about 80 minutes for each respondent, and a median total of 150 minutes if two interviews were conducted in a household. A substantial proportion of the interview in every wave is devoted to questions about income and assets, topics which respondents generally find more burdensome than questions about topics such as health or attitudes. In addition to the biennial interviews, by 2008 sample members had been asked to participate in an average of 3.11 supplemental data collections, which have included mail questionnaires sent in the years between core data collections, questionnaires left with respondents at the end of an interview, and data collections via the internet.

The HRS data collections have used a mixed-mode design. The baseline interviews with the original HRS sample, the older members of the AHEAD sample, and the CODA and WB samples, and follow-up interviews with respondents age 80 and older and with more than half of all respondents under age 80 starting in 2004, have mostly been conducted face-to-face, while the baseline interviews for the AHEAD cohorts of 1914–1923 and follow-up interviews with those members of all samples who were under age 80 were mostly been conducted by telephone through 2002.

The procedure with respect to incentives that was followed in follow-up data collections from 1993 through 1998 was as follows. A few days or weeks prior to their first contact with the sample members at each wave, the interviewers mailed them a letter reminding them of their prior participation and saying that they would be calling them soon to arrange a time for their next interview. Included with the letter was a University of Michigan check for $20 – or two checks, each for $20, if the letter was to both members of a couple. The procedures were modified at the end of the field periods for early waves of HRS. In particular, there was also a nonresponse study at the end of the baseline wave for the HRS sample in 1992: Those selected for this procedure were sent a Federal Express packet with a one-page letter explaining the importance of their participation and offering either

$50 or $100 per person ($100 or $200 per couple) to those agreeing (Juster and Suzman 1995; Lengacher et al. 1995). About 5% (612) of the baseline respondents were part of the nonresponse study.

2.2. *Incentive Experiment*

Prior to the start of data collection in 2000, there was concern that the value of the $20 that had been given to sample members since 1993 had fallen with inflation, and consideration was given to increasing it, with the expectation that this would increase the response rate and reduce the number of calls that interviewers would have to make and therefore reduce interviewing costs. Before instituting this change, it was considered prudent to do an experiment to test the validity of those expectations.

The probability of responding to a request for an interview is affected by a wide range of factors, including characteristics of the survey design, of the interviewers, of the social environment, and of the sample members (see, for example, Groves and Couper 1998). For purposes of the incentive experiment, it was considered important to distinguish between noninterviews due to refusal by the sample member, and those primarily due to cognitive or physical health limitations. The former are more likely than the latter to be reduced by external motivating factors such as higher incentives.

For this reason, the households in which interviews were sought in the 2000 wave were divided into four strata: 1) those in which the sample person (or at least one person in a couple) reported that his or her health was much worse at the time of the 1998 interview than it was two years previously ($n = 1{,}898$ sample members in 1,247 households); 2) those that were not in the first stratum but in which a proxy did the interview for the sample person (or for at least one person in a couple) ($n = 2{,}592$ sample members in 1,512 households); 3) those that were not in the first two strata but in which the sample person (or at least one person in a couple) was eligible but not interviewed in 1998 ($n = 2{,}622$ sample members in 1,651 households); and 4) all other households (16,569 living sample members in 11,268 households). If these strata definitions are applied to previous waves of HRS data collections, those in Strata 1 and 2 were considerably more likely to have died by the following wave than those in Stratum 4 (15.9% of those in Stratum 1 and 5.8% of those in Stratum 2, vs. 3.1% of those in Stratum 4 and 4.2% of those in Stratum 3). Differences between strata are even more striking with respect to response rates: more than half of those in Stratum 3 (55.4%) also failed to be interviewed (self or proxy), compared to 5.8% of those in Stratum 4; 8.3% of those in Stratum 1; and 11.6% of those in Stratum 2. Given these outcomes in these earlier waves, we expected that those in Stratum 3 would be considerably less likely to participate than other sample members, but we also expected that their attrition would be reduced by offering them higher incentives. On the other hand, those in Strata 1 and 2 were singled out because their circumstances (generally poor physical and/or cognitive health) often made their participation more difficult and therefore we expected that they would be less sensitive to changes in incentives.

Households within each of the four strata were randomly assigned to one of three treatments. Households with a total of about 300 eligible sample members were randomly selected from each of the first three strata, and with about 600 eligible sample members from Stratum 4, to receive the $30 (per sample member) treatment; like numbers were

randomly assigned to the $50 treatment; and the remaining households, with a total of 20,663 sample members, were assigned to the standard $20 treatment. Each sample member in a household, including new spouses and partners reported in the 2000 wave, received the same incentive.

3. Effects on Response Rates

3.1. Baseline Response Rates

Of the 15,497 individuals selected from the household screening in 1992, interviews were obtained with 12,654 respondents, for an overall response rate of 81.4% (multiplying the 99.6% screening rate by the 81.7% response rate among identified individuals). The strategy of offering reluctant respondents a large financial bonus for participation (as described in Subsection 2.1) increased the baseline response rate by about 4 percentage points. For the AHEAD sample, to whom the baseline interview was administered in 1993–1994, interviews were obtained with 8,222 respondents, for a response rate of 80.4%. The response rate for the CODA sample was 72.4%, and that for the WB sample was 69.9%. One likely reason for the lower response rate of the WB cohort is that that sample was generated in 1992, as part of the same screening operation that generated the HRS. There was little or no contact between the screening operation and the introductory letter inviting their participation in the study in 1998, so considerable effort was required to locate those who had moved and a substantial proportion could not be located. The CODA sample was selected from the Medicare Enrollment Database, and a substantial proportion of these as well could not be located. The differences may also be a reflection of a general trend, observed across many cross-sectional surveys, toward lower response rates (for example, see de Heer 1999; Atrostic et al. 2001; de Leeuw and de Heer 2002; Curtin et al. 2005).

3.2. Follow-up Wave Response Rates

Given that the incentive experiment was carried out in a follow-up wave, the response rates at each follow-up wave are more relevant to the evaluation of the incentive experiment than the baseline response rates. Table 1 displays the numbers of households and individual sample members, and the respective response rates, for each follow-up (nonbaseline) wave from 1994 through 2008. The proportion of sample members with or for whom interviews were obtained, and also the proportion of households with at least one interview, declined from about 90% to about 80% as the proportion of permanent refusals increased from 0% to 11% of the sample cases. The gradual decline in response rates occurred for both single and married sample members. If either member of a couple asks to be removed from the study, the practice has been to remove both, with the consequence that there has been a substantial increase, from 9% to 19%, in the proportion of couples in which neither member was interviewed.

The design yields three types of individuals with whom interviews are sought at each follow-up wave, and very different response rates have been achieved for those three groups: first, and by far the most numerous, are those who participated in the preceding wave; second are those who were eligible at the preceding interview but did not

Table 1. Sample sizes and response rates in follow-up waves from 1994–2008

Year	Number of households	Proportion of households with at least one interview	Number of individuals	Response rate for individuals	Reinterview rate
1994	7,908	91.4%	13,009	89.1%	91.8%
1995–96	14,457	90.2%	21,377	88.9%	93.2%
1998	14,128	88.0%	20,522	86.7%	93.9%
2000	16,787	86.7%	24,468	85.5%	93.9%
2002	16,177	85.6%	23,294	84.4%	94.7%
2004	15,237	83.5%	21,953	82.1%	95.3%
2006	14,637	81.7%	20,848	80.2%	95.4%
2008	13,928	81.0%	19,654	79.1%	96.1%
Total	113,259	85.7%	165,125	84.3%	94.3%

Note: All of the 1994 sample cases are from the original HRS sample. The 1995–1996 and 1998 samples combine the AHEAD and HRS samples. Starting with the 2000 follow-up wave, the WB and CODA samples are also included.

participate; and third are new spouses and partners who entered the sample because of their marriage to (or partnership with) a cohort-eligible sample member. The average response rate for the modal group (sample members who were interviewed, self or proxy, at the previous wave and therefore referred to as the "reinterview" rate) has been 94%. (Note that in calculating response rates for nonbaseline data collections, deceased as well as living sample members are included in the denominator; and in the numerator, interviews are counted whether conducted with a proxy informant or with the sample member, including interviews with proxies for deceased sample members.) The last column of Table 1 shows that the reinterview rate has slowly but monotonically increased across waves, from about 92% in 1994 to 96% in 2008. This increase has been observed in each of the four samples. As estimated from a probit model for each sample, the reinterview rate has increased by 0.5 to 0.6% per wave, and the linear trend (in probits) is statistically significant for each sample.

The response rates for those who were eligible at the previous wave but did not participate (referred to as "recontact" rates) have consistently been much lower than the reinterview rates and also much more variable, reflecting at least in part the amount of effort that was put into obtaining them. At the second wave of HRS, very little effort was made and this resulted in a recontact rate of only 8%. For the AHEAD, WB, and CODA samples, the recontact rates at their first follow-up waves were much higher (about 50%). With that exception, the recontact rates have not shown a systematic change across waves. Finally, the response rates for the small number of new spouses and partners reported at each follow-up wave vary widely and with no obvious pattern.

Across all waves through 2008, 9% of interviews for living sample members have been done with proxy informants, 64% of whom have been the sample member's spouse or partner. These proportions vary strongly with age: among those aged 60 or less, 6% of interviews have been with proxies, and over 90% of those were a spouse or partner, whereas among those over age 80, 18% of interviews have been with proxies and only 24% of those were a spouse or partner. Within age groups, the proportion of interviews

done by proxy has stayed fairly constant across waves, with no evidence of either an increasing or decreasing trend. About 95% of proxy informants for sample members who were aged 70 or younger and were married or living with a partner were spouses or partners. This drops to about 90% of those aged 71–80, and to 72% of those aged 81 and older, presumably reflecting the declining health of both members of older couples.

The proportion of follow-up interviews done face-to-face, rather than by telephone, has increased from 7% in the 1994 wave to over half of the interviews starting in 2004.

3.3. Immediate Effects of the Incentive Experiment

The response rates for those offered each of the three incentive levels and in each of the four strata are shown in Table 2. Tests of statistical significance shown were obtained from probit regression analyses, using the svy command in Stata to take account of the complex sample design. (The weights used for this and for all other analyses reported in this article were based on the household level weights for the 1998 wave, which take account of differences in the sampling rate, and are post-stratified to match Current Population Survey estimates of the distribution of housing units by the age, gender, marital status, and race of its members; for details, see http://hrsonline.isr.umich.edu/sitedocs/wghtdoc.pdf. The 1998 weights were used because the random assignments to incentive levels were based on status in 1998; and household rather than respondent weights were used because respondent weights are assigned values of 0 for nonrespondents, and thus to most of those in Stratum 3. To take account of the distinction between housing units with coupled vs. uncoupled sample members, the household weight was divided by two for coupled individuals.)

The comparison of a full model, which included six terms to take account of interactions between incentive levels and strata, with an additive (noninteractive) model led to the rejection of the additive model: $F(6,47) = 4.08$, $p = .0023$. Because the incentive effects differ from stratum to stratum, overall incentive effects, and overall differences between

Table 2. Response rates in 2000, by incentive group and by stratum

	Incentive		
Stratum	$20	$30	$50
1) Deteriorating health	93.9%	94.0% (ns)	96.5% (ns/ns)
	1,297	300	301
2) Proxy informant	91.7%	88.4% (ns)	94.0% (ns/ns)
	1,991	300	301
3) Noninterview	37.7%	39.5% (ns)	47.2% (*/ns)
	2,014	305	303
4) Everyone else	94.1%	97.5% (*)	98.4% (***/ns)
	15,361	606	602

Note: Sample sizes are given below the response rate achieved for each cell. Significance levels, as indicated in parentheses, test the difference in response rates between incentive levels: in the $30 column, the comparison of the $20 and $30 incentives; in the $50 column, first the comparison of the $20 and $50 incentives, second the comparison of the $30 and $50 incentives. The response rate estimates and significance levels shown in this and the following tables take the complex sample design and the weights into account. The significance levels (two-tailed) shown in this and the following tables are as follows: ns: $p > .05$ *, $p < .05$ **, $p < .01$ ***, $p < .001$.

strata, are not meaningful. The weighted response rate ranged from a low of 37.7% among those in Stratum 3 (nonrespondents in 1998) given the $20 incentive to a high of 98.4% among those in Stratum 4 (the residual group) given the $50 incentive. Within each of the four strata, those given $50 consistently had a higher response rate than those given $20, but only in Strata 3 and 4 is this difference statistically significant. In three of the strata, the response rate of those given the $30 incentive is higher than those given $20 and lower than those given $50, but the only statistically significant difference is for sample members in Stratum 4, among whom the response rate of those given $30 was significantly higher than that of those given $20. At each incentive level, those in Stratum 4 had the highest response rate, those in Stratum 3 the lowest, and those in Strata 1 and 2 (those with deteriorating health and those for whom proxy interviews were done in 1998) had response rates somewhat lower than those in Stratum 4. The lack of statistical significance for some of these comparisons, especially in Strata 1 – 3, may reflect in part the relatively low power afforded by the rather small sample sizes.

The pattern of results is consistent with the expectations that motivated the designation of four strata in the design of the incentive experiment. It was expected that those in Strata 1 and 2 would not be very sensitive to the size of the incentive, since the selection criteria for these strata (that is, much worse health, or a proxy informant, for at least one household sample member) were chosen to over-represent sample members for whom participation would be difficult because of physical or cognitive limitations; and indeed, Table 2 shows only small, statistically nonsignificant differences between incentive levels. Those in Stratum 3, on the other hand, were expected to be more strongly influenced by an increase in the incentive, since the refusal (by at least one sample member in the household) often reflects lack of motivation to participate, a lack that previous studies have shown may be overcome by higher incentives; and Table 2 shows that those in this stratum had a response rate 10 percentage points higher if given an incentive of $50 rather than $20. Finally, the rest of the sample after removing those in these special circumstances showed significantly higher response rates when given either $30 or $50 rather than the standard $20 incentives.

3.4. Possible Confounding Factors

To examine the possibility that the incentive effects described in the previous section are confounded by extraneous factors, additional analyses were carried out. First, as mentioned in Subsection 2.1, a nonresponse study was done at the baseline wave for the HRS sample: sample cases that had not responded to frequent requests for interviews were offered considerably higher incentives ($50 or $100). Not surprisingly, the sample members who only responded after getting that offer turned out to be more likely to refuse to participate in future waves, and so were more likely than other sample members to be in Stratum 3 (that is, nonrespondents in 1998: 21% vs. 11%). To test whether these cases influenced the estimates of the effects of higher incentives, the response rates within each of the four strata and at each of the three incentive levels were obtained after dropping these cases. None of these response rates differed by as much as one percentage point from the corresponding rate shown in Table 2. Furthermore, interaction terms between incentive levels and inclusion in the baseline nonresponse study were included in probit regressions, and none approached statistical significance.

A second concern is that some of the sample lines included in the incentive experiment based on their status in the 1998 wave had died by the time of the 2000 wave. As expected, those in Stratum 1 (deteriorating health) and in Stratum 2 (proxy interviews), and to a lesser extent those in Stratum 3 (nonrespondents), were more likely to have died than those in Stratum 4 (the residuals); but also as expected, mortality was not significantly related to the incentive level. A probit regression, predicting mortality from dummy variables for the size of the incentive, the strata, and the interactions between incentives and strata, showed that incentive size had no significant "effect" on mortality, and that there were no statistically significant interaction effects.

Third, as in every wave about 9% of the core interviews in the 2000 wave were done by proxy informants rather than by the sample individual him- or herself, and as expected this was much more likely for those in Stratum 2 (those with proxy interviews in 1998: 42%) than for those in Stratum 4 (the residual category: 4%), with those in Stratum 1 (deteriorating health: 19%) and Stratum 3 (noninterviews: 7%) also having elevated proxy rates. As with mortality, the choice to have a proxy rather than the sample person do the interview was not significantly related to the incentive level. A probit regression, predicting proxy vs. self interview among living sample members for whom a core interview was done, from dummy variables for the size of the incentive, the strata, and the interactions between incentives and strata, showed that incentive size had no statistically significant effect on the proxy/self decision, and that there were no statistically significant interaction effects.

Finally, the possibility was considered that the incentive effects are mediated by the interviewers. There were a total of 149 interviewers, and the response rates for sample lines initially assigned to individual interviewers ranged from 0% to 100% (69% to 99% among interviewers assigned at least 20 lines). Since interviewers are assigned sample lines within their geographic locales, this variation in response rates confounds differences in the characteristics of those locales and of the people living in them with any real differences between interviewers. The incentive treatments, moreover, were assigned randomly within strata, and there is no statistically significant relationship between interviewers and the incentives offered to households assigned to those interviewers. Therefore, it is not surprising that the pattern of response rates, after controlling on interviewer (that is, by including dummy variables for 148 of the 149 interviewers) is very similar to what is shown in Table 2.

3.5. Effects on Response Rates in Later Waves

To examine the longer term effects of the magnitude of the incentive provided in a given wave, probit regressions were estimated on the entire eligible sample at each of four subsequent waves (2002, 2004, 2006, and 2008), controlling on strata. New spouses and partners first reported after the 2000 wave are excluded from these analyses, since they did not receive the differential incentives given in 2000. Also, the number of eligible sample members declined across waves due to mortality (deceased individuals are dropped from these analyses the wave after a proxy interview for them has been completed). To provide context for interpreting the response rates in subsequent waves, it is relevant that starting in 2002, all respondents were offered a $40 incentive: that is, for most sample members the

incentive was doubled from the previous level of $20, but for those given the $30 incentive in 2000 the increase was only $10 (33%), and for those given $50 in 2000, the incentive actually decreased by $10 (−20%). (Because the incentive increase to $40 in the 2002 wave was not done as an experiment, it is not possible to estimate the effect of the increase on the response rate, since other changes could also have occurred. It is suggestive, however, that the reinterview rates in Table 1 are higher in the 2002 and each subsequent wave – 95–96% – than in any of the previous waves – 92–94%.)

Unlike what was found in the 2000 wave, no statistically significant interactions were found between incentive level and stratum: that is, in each of the four waves, 2002 through 2008, the incentives offered in the 2000 wave had approximately the same impact on the response rates in each stratum. The second through fifth columns of Table 3 display the estimated response rates for sample members in each of the four strata (estimated at the $20 incentive level) and at each of the three incentive levels (estimated for Stratum 4). These response rate estimates are derived from probit regressions predicting participation in each of the four waves. Those given $50 in 2000 had significantly higher response rates in all four waves than those given $20 (despite their experience of a decrease in the incentive, to $40 starting in 2002). The overall response rates of those given $30 incentives in 2000 were generally slightly higher than those given $20 and somewhat lower than those given $50; only the difference between $30 and $50 in 2004 is statistically significant. (Again, the power for detecting statistical significance for these and most of the other comparisons made in this article is limited because of somewhat small sample sizes of the incentive groups in Strata 1 to 3.) The final column of Table 3 shows the coefficients from an OLS regression analysis predicting the proportion of the waves in which respondents were interviewed, taking into account their eligibility at each wave. (The same pattern and significance levels are obtained from Poisson regressions predicting the number of noninterviews for each sample member across the four waves.) The average response rate across the four waves was 3.3 percentage points higher for those given the $50 incentive in 2000 than for those given $20 – a difference that is both statistically and practically significant. Since the overall nonresponse rate with the $20 incentive is in the range of 8–10%, a decrease of over 3 percentage points would be an important improvement with respect to the long-term viability of the sample, reducing long-term attrition and likely the magnitude of selection biases. (The magnitude of nonresponse biases depends on two factors: 1) differences in the distributions of variables between respondents and nonrespondents; and 2) the proportion of nonrespondents. The first of these factors has not been examined in this article, so nothing definitive can be concluded about the existence or magnitude of nonresponse biases, but it is certainly plausible to suspect that nonrespondents differ from respondents on variables of interest to some analysts.) Those given $30 had a response rate that is midway between those given the lower and higher incentives, but neither difference is statistically significant. To summarize the main point of this analysis: if the incentive had been increased to $50 for all sample members in the 2000 wave, and otherwise the survey incentives were unchanged (that is, kept at $40 for all sample members starting in 2002), it is estimated that the number of interviews across the five waves, 2000 through 2008, would have increased by almost 4,000. The point estimate, 3,911 additional interviews, was obtained by taking the number of sample members in a given wave and stratum times the estimated response rate

Table 3. Response rates in 2002–2008 predicted by incentive group and by stratum

	Year				Average
	2002	2004	2006	2008	2002–2008
Sample size	22,374	20,889	19,668	18,395	22,374
Stratum (at $20 incentive)					
1) Deteriorating health	92.3% (ns)	90.6% (ns)	87.5% (ns)	86.5% (ns)	91.5% (ns)
2) Proxy informant	87.8% (***)	84.6% (***)	80.9% (***)	77.7% (***)	84.8% (***)
3) Noninterview	44.3% (***)	44.0% (***)	40.5% (***)	39.2% (***)	45.6% (***)
4) Everyone else	92.7%	90.4%	88.7%	87.9%	90.7%
Incentive (in Stratum 4)					
$20	92.7%	90.4%	88.7%	87.9%	90.7%
$30	94.3% (ns)	89.5% (ns)	90.4% (ns)	89.9% (ns)	92.0% (ns)
$50	94.8% (*/ns)	92.7% (**/**)	92.0% (**/ns)	91.6% (**/ns)	94.1% (**/ns)

Note: Estimates assume additivity of incentive and stratum effects. The estimates in the second through fifth columns are derived from probit regressions; those in the last column are based on an OLS regression predicting the sample member's response rate across the waves at which each sample member was eligible. For the strata, the significance (indicated in parentheses) is with respect to the difference in the response rate for those in the given stratum vs. those in Stratum 4. For the incentives, the significance is with respect to the difference in the response rate for those offered the given incentive vs. those offered $20. The second significance level (after the slash) in each cell of the last row refers to a test of the equivalence of the response rates for the $30 and $50 incentives.

at that wave and stratum for those in the $50 incentive group, and subtracting the corresponding number assuming all would have had the response rate observed among those in the $20 incentive group; and summing this difference across all four strata and all five waves. To put this in perspective, the number of sample members who were included in the 2000 incentive experiment for whom interviews were sought in the 2008 wave was 16,934, and interviews were obtained with or for 15,105 of them.

4. Effects on Data Quality

Apart from the effects of incentives on response rates, another consideration is the impact of incentives on the quality of the data provided by the respondents. The impacts of increasing the incentive size on two indicators of data quality are examined in this section.

4.1. Differences in Item Missing Data Rates Associated With Higher Incentives

To evaluate the effects, positive or negative, of higher incentives on item missing data, it is first necessary to take into account that a large proportion of the questions in the HRS questionnaire are not asked of every respondent. Many questions are skipped based on the answers to prior questions: for example, questions about work are not asked of those who report they are not working, and questions about medical care for specific health conditions are not asked of those who do not report having those conditions. Moreover, while the proportion of questions that are asked but not answered (i.e., the respondent either refuses to answer or says that he/she does not know the answer) is generally low, it varies across questions. To take account of the specific subset of questions that each respondent was asked, as well as differences in the missing data rates for those questions, the following statistics were calculated for each respondent. The first is the number of questions to which a respondent (i) said "don't know" (DK) minus the expected number, where the latter is calculated as the proportions of DK responses summed across the subset of questions asked of a particular respondent, i:

$$D_{Di} = \sum_{k=1}^{K} \left\{ M_{Dik} - \frac{\sum_{j=1}^{n} M_{Djk}}{\sum_{j=1}^{n} S_{jk}} S_{ik} \right\} \tag{1}$$

where K = 2,102 is the number of questions examined; M_{Dik} and M_{Djk} are indicator variables for DK responses (e.g., $M_{Dik} = 1$ if focal respondent i gave a DK response to item k, $= 0$ otherwise); $n = 19{,}580$ is the total number of respondents, and S_{ik} and S_{jk} are indicator variables for whether or not respondent i or j was asked item k (e.g., $S_{jk} = 1$ if respondent j was asked item k, $= 0$ otherwise). The second statistic is parallel to the first, but assesses the tendency to refuse to answer questions:

$$D_{Ri} = \sum_{k=1}^{K} \left\{ M_{Rik} - \frac{\sum_{j=1}^{n} M_{Rjk}}{\sum_{j=1}^{n} S_{jk}} S_{ik} \right\} \tag{2}$$

Questions that were asked of fewer than 100 respondents, and those to which no respondent gave either a "Don't Know" or a "Refusal" response, were excluded. On average, individual respondents were asked 313 of these questions in the 2000 wave.

Regression models were estimated, predicting each of the indices defined by Expressions 1 and 2, from incentive level and stratum. (The set of terms for the interactions between incentive and strata was found to be statistically nonsignificant with respect to both indices.) The estimated values of the indices defined by Expressions 1 and 2 are shown in Table 4 for respondents in each of the four strata and at each of the three incentive levels. Some of the differences between strata are highly significant: respondents in Strata 1, 2, and 3 said that they did not know the answers to 1.5 to 2.2 more questions than did those in Stratum 4; and those in Stratum 3 (i.e., those who refused to be interviewed at all at the previous wave) refused to answer an average of 2.7 more questions than did those in the other strata. The differences between the incentive groups are much smaller than those between strata, and most are not statistically significant, but the tendency is for those given higher incentives to have less missing data than those given the standard $20 incentive. Specifically, those given $30 said "don't know" to fewer items than those given $20. Also, there was a nonsignificant tendency for those given $50 to refuse to answer fewer items than those given $20 or $30.

In addition to the indicators defined by Expressions 1 and 2, indicators defined as the ratio of the observed to the expected counts of "don't know" and "refusal" were created, but the patterns of regression coefficients for strata and incentive levels are similar to those shown in Table 4, with one exception: the difference in the ratio of refusals to the expected count between those given $50 and those given $20 was now statistically significant. Additional analyses showed that the patterns did not change after truncating extreme values on the indicators. Another set of regression models added the corresponding index of missing data from the 1998 wave. This analysis was restricted to sample members who

Table 4. Predict indicators of "don't know" and "refusal" responses by incentive group and stratum

	Observed − expected	
	Don't knows	Refusals
Stratum (at $20 incentive)		
1) Deteriorating health	2.23 (***)	−.08 (ns)
2) Proxy informant	1.51 (***)	.21 (*)
3) Noninterview	2.22 (***)	2.74 (***)
4) Everyone else	−.01	−.04
Incentive (in Stratum 4)		
$20	−.01	−.04
$30	−.67 (**)	−.01 (ns)
$50	−.15 (ns/ns)	−.36 (ns/ns)

Note: The entries are indices of item missing data responses estimated assuming additivity of incentive and stratum effects derived from OLS regressions. For the strata, the significance levels (in parentheses) are with respect to the difference in the frequency of missing data for those in the given stratum vs. those in Stratum 4. For the incentives, the (first) significance is with respect to the difference in the frequency of missing data for those offered the given incentive vs. those offered $20. The second significance level (after the slash) in each cell of the last row refers to a test of the equivalence of the frequency of missing data for the $30 and $50 incentives.

did self interviews in both waves, which eliminated most of the sample members in Strata 2 and 3 and a large proportion of those in Stratum 1, so only the estimates for Stratum 4 are meaningful. The same patterns of estimates and their statistical significance were found as those shown in Table 4.

4.2. *Effects on Data Quality: Differences in Propensity to Say "No" Associated With Higher Incentives*

As noted earlier, Zagorsky and Rhoton (2008) reported that sample members in a long-running panel study who refused an interview at one wave but were interviewed at the next wave were asked more questions during that interview if they were offered an incentive than if they (following past practice) received no incentive. They attributed this difference to the incentive increasing the motivation of these respondents to report activities or relationships even though they realized that those affirmative reports would result in their being asked additional questions. In the HRS interview, there are many such questions: for example, if a respondent reports a medical condition, they are asked questions about treatment for and severity of each such condition; and if they report receiving any income from a particular source (e.g., an annuity or stock holdings), they are asked about the amount received, the frequency, and so on. Many have learned this consequence of affirmative responses, and those with low motivation to participate may be more likely to reduce the interview length by fudging their answers to the lead-in question. They can avoid the follow-up questions either by an outright "no" response to the lead-in question, or by declining to answer it (that is, by saying that they "don't know" or by refusing any response). To assess the impact of incentive size on the propensity to avoid follow-up questions, a similar strategy to that described in Subsection 4.1 for examining the propensity toward missing data, but now restricting the questions to those with answer categories of "yes" (including qualified "yeses") and "no." Indices were defined in the same way as shown in Expressions 1 and 2, first looking only at the "no" responses, then including the DKs and refusals along with the "no" responses. This was done separately for each of twelve major topic areas within the survey, and overall across the entire interview: a total of 318 items (excluding items asked of fewer than 100 respondents). The pattern is illustrated in Table 5, using an index that combined missing data (DK or refusal) responses with "no" responses, across all sections of the interview. The only significant effect of incentives is observed for those with deteriorating health (Stratum 1). At each incentive level, those in Stratum 1 gave fewer "no" responses than those in any of the other three strata, and this is primarily because they were less likely to deny having various health conditions – a reasonable pattern, given that they had reported their health to be much worse in 1998. Those in this stratum who were given the $50 incentive, however, gave an average of 1.57 more negative responses than those in the same stratum that were given $20, suggesting that the higher incentive could have a negative impact on data quality for those in poor health. Table 5 also shows that among those who refused to be interviewed at the previous wave (Stratum 3), those given $50 incentives gave 0.9 fewer negative responses than those given $20. That difference is not statistically significant, but the pattern is consistent with Zagorsky and Rhoton's (2008) finding.

Table 5. Predict indicators of tendency to give responses that avoid follow-up questions, by incentive group and stratum

	Incentive provided		
Stratum	$20	$30	$50
1) Deteriorating health	−2.55	−2.14 (ns)	−.98 (**/ns)
2) Proxy informant	−.45	−.16 (ns)	−.31 (ns/ns)
3) Noninterview	.79	.51 (ns)	−.11 (ns/ns)
4) Everyone else	.08	.14 (ns)	−.14 (ns/ns)

Notes: Entries are averages of the counts of "yes/no" items to which each respondent gave "no," "don't know," or "refusal" responses, minus the expected number of items with that type of response for that same subset of items. The significance levels shown in the cells in the third column, and the first one in the fourth column, apply to tests of the difference between those given the $30 or $50 incentive and those in the same stratum given the $20 incentive. The second significance level (after the slash) in each cell of the last column refers to a test of the equivalence of the response rates for the $30 and $50 incentives.

5. Discussion

This article has looked at a range of possible consequences of changing the size of incentive offered to sample members in a panel study. The evidence from an experimental study carried out during one wave of HRS indicates that the response rate increases with the size of the incentive. The increase was especially great among sample members who had been nonrespondents at the previous wave. This pattern is consistent with one reported by Trussell and Lavrakas (2004) when they had mailed incentives ranging from $0 to $10 to respondents and nonrespondents on an initial telephone survey: the larger the incentive, the higher the rate of return of a mail questionnaire, and the slope was greater for the initial nonrespondents than for the initial respondents. The pattern is also consistent with the finding reported by Zagorsky and Rhoton (2008) that offering an incentive increased the response rate of those who had been nonrespondents at the previous wave, thereby stabilizing or even reducing the gradual attrition experienced over the initial 20 or so waves when no incentives were offered.

The results from the experimental study also indicate that an increased incentive has little or no effect on sample members who have difficulty with the respondent task, because of physical or cognitive limitations, as indicated either by their reports of a serious decline in their health at the previous wave, or by their previous wave interview having been done by a proxy.

The implication is that panel attrition can be reduced by increasing the incentive offered to the sample members. There is also the implication that a cost-effective strategy would be to tailor incentives based on what is known about sample members. Incentives, or higher incentives, could, for example, be offered only to those who are reluctant to be interviewed, or offered to all sample members except those are not expected to be influenced by incentives. Such strategies, however, have often been viewed unfavorably out of concerns for fairness and of possible negative impact on future waves if sample members were to become aware of the practice. This issue is discussed in Singer et al. (1999a), but the authors did not find any statistically significant support for the hypothesis in an experimental study.

Various hypotheses have been offered about incentive effects, with different implications for survey practice. One possibility is that higher incentives increase the

credibility and perceived importance of the study; if so, we would expect that those offered higher incentives at one wave have higher levels of participation at later waves, even if the incentive level at those later waves is equal for all sample members. Another possibility is that participation at one wave, however it is induced, increases the probability of participation at the next wave (and, by extension, for all future waves), and again the implication is that higher incentives at one wave would lead to higher participation at later waves by virtue of increasing the response rate at the immediate wave.

Interviewer behavior may be affected by the size of the incentive (e.g., they may feel more confident about persuading reluctant sample members to be interviewed if they know that he or she received a large incentive). The HRS interviewers were not informed of the incentive experiment, but it is possible that some sample members mentioned that they had received $30 or $50 instead of the then standard $20. When dummy variables for individual interviewers were included in the probit regressions, however, there was no diminution in the estimated effects of higher incentives on response rates.

In designing the experimental study, there was concern was that there could be a negative impact on response rates if incentive size were to be increased at one wave but reduced at later waves. If higher incentives in the 2000 wave were to have raised the sample members' expectations about incentives in future waves, we would have expected the respondents given $50 in 2000 to have lower response rates in later waves when they were given only $40, but this is the opposite of what we found. The sample members who were given $50 incentives in the 2000 wave tended to have higher response rates in later waves despite the fact that their incentive was reduced to $40. This is consistent with a finding reported by Lengacher et al. (1995) based on the nonresponse study done at the end of the baseline data collection for HRS: sample members who were given large incentives ($50 or $100 per respondent) at Wave 1 had about the same response rate at Wave 2 as reluctant respondents who were given the standard incentive (which was $10 at that wave). The implication is that providing a larger incentive at any wave of a panel study, even one that is not repeated at later waves, is effective in reducing panel attrition at that wave but also at future waves. This pattern is inconsistent with Hansen's (1980) self-perception hypothesis, which would predict that those respondents who receive a larger incentive at a given wave are less likely to perceive themselves as good survey respondents ("I just did it for the money") than those who receive a smaller incentive, and therefore are less likely to continue their participation at later waves. The pattern is, however, consistent with Dillman's (1991) social exchange hypothesis, which predicts that those who perceive the benefits – extrinsic as well as intrinsic – of being a panel member as outweighing the costs of doing so are more likely to participate.

Another concern is with respect to the quality of the data, including item missing data rates and the accuracy of the answers that are given. Offering higher incentives could induce some sample members to participate in order to justify their receipt of the incentive, but might not increase their motivation to engage in the task and provide thoughtful answers. On the other hand, higher incentives could motivate some respondents to take their role more seriously, and therefore exert more effort in listening to the questions and processing their responses. Analysis of the data from the 2000 wave of HRS showed some statistically significant decreases in item nonresponse rates among those given higher incentives. There was also a tendency, albeit not always statistically

significant, that is consistent with Zagorsky and Rhoton's (2008) finding that an incentive increased the number of questions answered by respondents who had refused to participate at the previous wave of data collection.

There are some obvious limitations on the generalizability of the findings from this study. The target population was limited by age: only those born in 1947 or before, plus their spouses or partners, were eligible. Moreover, despite the relatively large overall sample size, the numbers in three of the four strata, and the numbers given the larger incentives within each stratum, were much smaller, thereby limiting the power, and therefore numerous patterns were noted that are suggestive but not statistically significant. Finally, it is not known how sensitive the findings are to the specific history of respondent incentives offered over the course of the HRS, including the fact that the incentives are given unconditionally. In this regard, however, it is noteworthy that positive effects of the higher incentives observed at the 2000 wave persisted at and after the 2002 wave in which the standard incentive was increased from $20 to $40. Moreover, in evaluating the generalizability of the findings it is important to keep in mind that sample members in each of four strata were randomly assigned to the incentive treatments, so characteristics of the sample members and of their treatment in previous waves are, within the limits of sampling variability, unrelated to observed differences in outcomes within each of the strata.

The importance of experimental studies of this sort should not be underestimated. Retaining sample cases is generally deemed to be critical to the long-term viability of longitudinal studies, as evidenced by the widespread practice of offering incentives (reviewed in Laurie and Lynn 2009). Much still needs to be learned about the optimal allocation of survey budgets to incentives, which must be weighed against the costs of other survey design decisions such as mode of data collection, frequency and type of contact with sample members between waves, and interviewer training and incentives (see Castiglioni et al. 2008). Little is known about the total cost associated with increasing respondent incentives; it is possible that total costs may actually be reduced by optimal decisions about incentives, if higher incentives are associated with greater cooperativeness by sample members, thereby reducing the average number of calls interviewers need to make to complete interviews (Singer et al. 2000; Rodgers 2002).

Moreover, there is uncertainty about the advantages and disadvantages of tailoring the size of incentives to sample member characteristics, such as their history of participation in previous waves, their frequency of item missing data, and any factors that make their data more or less informative with respect to the objectives of the study. And finally, much remains to be learned about the effectiveness of any increases in response rates that may result from higher incentives on reducing biases in sample estimates of population characteristics.

6. References

Atrostic, B.K., Bates, N., Burt, G., and Silberstein, A. (2001). Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights. Journal of Official Statistics, 17, 209–226.

Castiglioni, L., Pforr, K., and Krieger, U. (2008). The Effects of Incentives on Response Rates and Panel Attrition: Results of a Controlled Experiment. Survey Research Methods, 2, 151–158.

Curtin, R., Presser, S., and Singer, E. (2005). Changes in Telephone Survey Nonresponse over the Past Quarter Century. Public Opinion Quarterly, 69, 87–98.

de Heer, W. (1999). International Response Trends: Results of an International Survey. Journal of Official Statistics, 15, 129–142.

de Leeuw, E. and de Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: Wiley, 41–54.

Dillman, D.A. (1991). The Design and Administration of Mail Surveys. Annual Review of Sociology, 17, 225–249.

Dillman, D.A. (2000). Mail and Internet Surveys: The Tailored Design Method. New York: Wiley.

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

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

Heberlein, T.A. and Baumgartner, R. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. American Sociological Review, 43, 447–462.

Jäckle, A. and Lynn, P. (2008). Respondent Incentives in a Multi-Mode Panel Survey: Cumulative Effects on Nonresponse and Bias. Survey Methodology, 34, 105–117.

James, T.L. (1997). Results of the Wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation. Proceedings of the American Statistical Association, Section on Survey Research Methods, 834–839.

Juster, F.T. and Suzman, R.M. (1995). An Overview of the Health and Retirement Study. Journal of Human Resources, 30(Suppl.), S7–S56.

Laurie, H. and Lynn, P. (2009). The Use of Respondent Incentives on Longitudinal Surveys. In Methodology of Longitudinal Surveys, P. Lynn (ed.). New York: Wiley, 205–233.

Lengacher, J.E., Sullivan, C.M., Couper, M.P., and Groves, R.M. (1995). Once Reluctant, Always Reluctant? Effects of Differential Incentives on Later Survey Participation in a Longitudinal Study. Proceedings of the American Statistical Association, Section on Survey Research Methods, 1029–1034.

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. Section on Survey Research Methods, 529–534.

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

Rodgers, W.L. (1996). Comparisons of Two Sampling Frames for Surveys of the Oldest Old. In Health Survey Research Methods Conference Proceedings, R.B. Warnecke (ed).

DHHS Publication No. (PHS)96-1013. Centers for Disease Control and Prevention, National Center for Health Statistics, Public Health Service, U.S. Department of Health and Human Services, 117-122.

Rodgers, W.L. (2002). Size of Incentive Effects in a Longitudinal Study. Proceedings of the American Statistical Association, Section on Survey Research Methods, 2930–2935.

Ryu, E., Couper, M.P., and Marans, R.W. (2005). Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error. International Journal of Public Opinion Research, 18, 89–106.

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley, 163–177.

Singer, E., Groves, R.M., and Corning, A.D. (1999a). Differential Incentives: Beliefs about Practices, Perceptions of Equity, and Effects on Survey Participation. Public Opinion Quarterly, 63, 251–260.

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

Singer, E., van Hoewyk, J., and Maher, M.P. (2000b). Experiments with Incentives in Telephone Surveys. Public Opinion Quarterly, 64, 171–188.

Soldo, B.J., Hurd, M.D., Rodgers, W.L., and Wallace, R.B. (1997). Asset and Health Dynamics Among the Oldest Old: An Overview of the AHEAD Study. The Journals of Gerontology, Series B, 52B(Special Issue), 1–20.

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

Yu, J. and Cooper, H. (1983). A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. Journal of Marketing Research, 20, 36–44.

Zagorsky, J.L. and Rhoton, P. (2008). The Effects of Promised Monetary Incentives on Attrition in a Long-Term Panel Survey. Public Opinion Quarterly, 72, 502–513.

Received April 2010
Revised February 2011