

SIZE OF INCENTIVE EFFECTS IN A LONGITUDINAL STUDY

Willard Rodgers

Survey Research Center, University of Michigan, Ann Arbor, MI 48106-1248

KEY WORDS: Panel attrition, Response rate

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 been found consistently and substantially to increase the response rate (Heberlein and Baumgarten, 1978; Yu and Cooper, 1983). A recent 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., 1999) 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. A recent review (Singer, 2002) indicates that little research has been done on the usefulness of incentives for maintaining high response rates in panel studies, or on the optimal size of such incentives. 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 increase incentives could prove to be a useful investment in the long-term viability of such studies.

This paper describes an experiment conducted as part of one wave of the Health and Retirement Study (HRS), which is a longitudinal survey of a nationally representative sample of persons who were born in 1947 or before. Interviews have been conducted biennially since 1992. The general practice, after wave 1, has been to include a check for \$20 along with an advance letter sent to each sample member prior to their being contacted, by telephone or face-to-face, by an interviewer. The experiment consisted of changing the amount of the incentive check for randomly selected sample members, with the objective of examining the effect of incentive size in the response rate, on the amount of effort required to finalize the case, and the quality of the data obtained from respondents. In this paper, we examine the effects of the size of the incentive on the response rate on the immedi-

ate wave, and also briefly examine the effects on the number of calls.

Methods

Design of HRS. HRS began in 1992 as a longitudinal survey of a narrow range of birth cohorts (those born between 1931 and 1941). The original HRS respondents have been reinterviewed every two years through 2002. In 1998, the HRS panel was merged with that of a parallel study, Asset and Health Dynamics among the Oldest Old (AHEAD), which followed a sample of those born before 1924 starting in 1993; and these two samples were supplemented with samples from the birth cohorts of 1924-30 (the "Children of Depression Age," or CODA sample) and 1942-47 (the "War Babies," or WB sample), making the overall HRS sample representative of the U.S. population age 51 and over in 1998. Details on the design of the HRS and AHEAD studies are provided elsewhere (Juster and Suzman, 1995; Soldo, Hurd, Rodgers, and Wallace, 1997); a brief summary of features of the design that are most relevant to the purpose of this paper is provided here.

The HRS data collections have used a mixed-mode design. The baseline interviews with the original HRS birth cohorts, the older AHEAD birth cohorts, and the CODA and WB cohorts, and follow-up interviews with sample members age 80 and older, have mostly been conducted face-to-face, while the baseline interviews for the AHEAD cohorts of 1914-23 and follow-up interviews with those members of all cohorts who were under age 80 have mostly been conducted by telephone. Approximately nine percent of the interviews have been conducted with proxy informants, generally because the sample person has physical or cognitive limitations that made their participation difficult or impossible, but occasionally also if a sample person refused to do an interview but was willing to let someone else answer the questions.

Sample members are 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 rarely, when a sample member or his or her spouse or partner explicitly asks to be permanently removed from the study, and those cases are kept in the denominator when response rates are calculated.

The standard procedure with respect to respondent incentives that has been followed in every data collection following the baseline inter-

view of the original HRS sample in 1992 is as follows.¹ A few days or weeks prior to their first contact with the sample members at each wave, the interviewers mail them a letter reminding them of their prior participation and saying that they will be calling them soon to arrange a time for their next interview. Included with the letter is a University of Michigan check for \$20 – or two checks, each for \$20, if the letter is to both members of a couple.

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 respondents since 1994 had fallen with inflation, and consideration was given to increasing it to \$50, 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.

All households that were eligible for interviews in 2000 were randomly assigned to one of three treatments. The eligible households were first divided into four strata: 1) those in which the sample person (or at least one person in a couple) reported the his or her health was much worse at the time of the 1998 interview than it was two years previously; 2) those not in the first stratum in which a proxy did the interview for the sample person (or, again, for at least one person in a couple); 3) those not in the first two strata in which the sample person (or at least one person in a couple) was not interviewed in 1998; and 4) all other eligible households. Households with a total of about 300 eligible sample members from each of the first three strata, and households with about 600 eligible sample members from the fourth stratum, were assigned to receive \$30; like numbers were assigned to receive \$50, and the remaining were assigned to receive the standard \$20.

¹ The procedure at Wave 1 of HRS in 1992 was more complex. Most single respondents were given \$10, and most married or partnered couples were given \$30, but interviewers were allowed considerable discretion especially toward the end of the data collection with respect to reluctant sample members. There was also an “endgame” at the end of that wave: Those selected for this procedure were sent a Federal Express packet with a one-page letter explaining the importance of their participation and offering \$100 per person, \$200 per couple to those agreeing (Juster & Suzman, 1995, p. S46). A similar endgame was instituted at the end of the third wave with the original HRS sample, in 1996.

Findings

Response rates and numbers of interviews. At the baseline HRS data collection in 1992, a total of 15,497 individuals were eligible for interviews. This total included persons identified in the household screening as born between 1931 and 1941, plus their spouses or partners regardless of year of birth. Of those identified in this way, interviews were obtained with 12,654 respondents, for an overall response rate of 81.7 percent. The “endgame” strategy of offering reluctant respondents a large financial bonus for participation increased the baseline response rate by about 4 percentage points. For the AHEAD sample, to whom the baseline interview was administered in 1993, interviews were obtained with 8,222 respondents, for a response rate of 80.4 percent. The response rate for the CODA sample was 72.4 percent, and that for the WB sample was 70.1 percent.²

In addition to the initial response rate at the time of the baseline data collection, the HRS 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, the response rate for those who participated in the preceding wave; second, the response rate for those who were eligible at the preceding interview but did not participate; and third, the response rate for new spouses who enter the sample because of their marriage to a cohort-eligible sample member, plus those who were married to a participating respondent but not interviewed in previous waves. The reinterview rates (i.e., the response rates for those who were interviewed at the previous wave) have all been in the range of 92 to 95 percent. For HRS and AHEAD, there are slight upward trends from the first wave of reinterviews (1994 and 1995/96, respectively) to 2000. Recontact response rates (i.e., the response rates for those who were eligible at the previous wave but did not participate) have consistently been much lower than the reinterview response 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 is reflected in a recontact rate of only 8 percent (the only cases who were recruited at wave 2 were the spouses and partners of those who were interviewed at wave 1). For the AHEAD, WB, and

² Reasons for the lower response rates of these latter cohorts likely have to do with peculiarities of the sample design, which are explained in a working paper version of this paper that is available from the author.

CODA samples, the wave 2 recontact rates were much higher (about 50 percent). The recontact rates have declined across later waves for both the HRS and AHEAD samples (to about 30 percent at wave 4 for both samples, and only 20 percent at wave 5 for the HRS sample). 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.

Results of 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 1. The overall response rate was 87.9%, and varied from a low of 38.3% among those in stratum 3, to 91.8% for those in stratum 2, to 94.1% of those in stratum 1, and 94.8% of those in stratum 4. Within each of the four strata, those given \$50 consistently had a higher response rate than those given \$20, and in all but stratum 2 this difference is statistically significant. The response rate for those given \$30 is generally, but not always (stratum 2 is again the exception) intermediate between the response rates for those given \$20 and \$50.

From Table 1, it appears that the response rate for those given \$30 is intermediate between the response rates for those given \$20 and \$50. To test whether the response rate is related in a linear manner to the response rate (or, more precisely, to the logit transformation of the response rate), a model was specified that included a linear (or log-linear) term for the size of the incentive, plus a dummy variable for those given \$30, and tested for evidence against the linear effect hypothesis by looking at the coefficient for the \$30 incentive variable. This model was estimated both for the entire sample (including dummy variables for the strata) and separately for each of the four strata. The hypothesis of a non-linear effect could not be rejected for the entire sample or for three of the four strata. The exception is stratum 2, for which Table 1 shows that the response rate among those given \$30 was actually somewhat lower than that for those given \$20.

To better understand the incentive effects, it is important to take more complete account of one aspect of the overall design and how that impacts the design of the incentive experiment. Eligible sample members may be either "coupled" (i.e., either married or living with a partner) or "uncoupled," and this may change from one wave to the next. For the incentive experiment, both members of a couple at the time of the 1998 data collection (or at the last wave in which at least one of them was interviewed) were assigned to the same incentive level for the 2000 wave, regardless of their marital status in 2000. If a respondent reported living with a new or different spouse or

partner in 2000, the new individual was given the same incentive as the old sample member. Therefore, for respondents who were coupled in 2000, the total incentive payment to their household was double the level assigned to them as individuals. The next analysis asks whether incentive level has different effects on coupled than on uncoupled respondents.

Table 2 shows the additive model (with terms for the level of incentive and for stratum, but not for their interaction) for each of three conditions. The first is predicting to the response rate for sample members who were neither married nor living with a partner at the time of the 2000 data collection. For coupled sample members, the overall response is broken into two steps: first, did at least one member of the couple complete an interview? and second, conditional on the participation of at least one member, did both complete an interview? In all three regressions, those receiving the \$50 incentive had a higher response rate than those offered \$20. There are differences between the models, though, with respect to the effectiveness of the \$30 incentive. For uncoupled respondents, and for the probability of getting at least one interview from couples, those receiving \$30 did not differ significantly from those receiving \$20, and the \$50 incentive yielded higher response rates than did \$30. In couples in which at least one member was interviewed, however, the \$30 incentive was at least effective as \$50 in increasing the probability that both members of the couple would be interviewed.

Differential incentive effects between subgroups. Attrition of a panel is important not only because of the reduction of sample size, but also because it may reduce the accuracy with which the remaining respondents represent the target population if some types of individuals are less likely to continue their participation than others. To examine the effectiveness of increased incentives in maintaining the representativeness of the respondents, dummy variables were added to the regression model for various characteristics of the sample members, including their gender (male vs. female), their race (Blacks vs. others), their ethnicity (Hispanic vs. others), and their birth cohort (those born in 1920 or before; 1921-30; 1931-40; and 1941 or later). There are suggestions of differences from these analyses (not shown): that increasing the incentive to \$50 may not have as much effect on the response rate of Hispanics as on that of non-Hispanics, and that the effectiveness of the higher incentive may be greater for younger than for older sample members. None of these interactions, however, is statistically significant.

Effect of incentives on number of interviewer contacts. Interviewers often have to make multiple

attempts to contact sample members before they succeed in conducting an interview or reaching the conclusion that the case is a final non-respondent for a given wave. The median number of distinct calls (after deleting call records to a household made within an hour of another contact) in the 2000 wave was 5, but the distribution is skewed, with a range from 1 to 110 and an interquartile range of 5 (from 3 to 8). The mean number of contacts was 6.98, with a standard deviation of 7.3. Most of these contacts are by telephone, but some are face-to-face, either because the case was assigned that mode or because the interviewer had difficulty making contact by telephone. Over half of the households required no face-to-face contacts, but the range extended to a maximum of 25 contacts, and the mean number was 0.48 with a standard deviation of 1.09. The average number of telephone contacts was 6.50, with a standard deviation of 7.06.

To determine whether increasing the incentive had any impact on the number of contacts required to finalize households, regression models were estimated, predicting to the number of contacts from the incentive level, the strata, and whether there were one or two sample members in the household. Providing the \$50 incentive to each sample member reduced the number of face-to-face contacts by .115, and the number of telephone contacts by .646, compared to those given \$20. Those given \$30 also required fewer contacts, but this is statistically significant only with respect to the number of face-to-face calls. Other coefficients indicate that 1.1 additional contacts were required to complete households with two sample members compared to households with just one sample member; that households in which at least one sample member was interviewed by proxy (stratum 2) required 1.2 more contacts than those in the "normal" group (stratum 4); and that households in which at least sample member was a non-respondent in the 1998 wave required 0.8 more face-to-face contacts, on average, and 3.8 more telephone contacts than those in the "normal" group.

The regression analyses just described for the entire 2000 sample were then repeated for each of the four strata separately to learn whether the higher incentives had differential effects on the number of contacts required to finalize these different groups of respondents. There are a few differences between strata that appear to be statistically significant: the \$50 incentive appears to be effective in reducing the number of telephone contacts for those in stratum 1, and the number of face-to-face contacts in strata 2 and 3, while \$30 and \$50 seem to be equally effective in reducing the number of telephone contacts in stratum 4.

The costs of increasing incentives. There is a direct cost associated with increasing the incentives given to respondents, and this must be weighed against the benefits to the study in terms of higher response rates and better maintenance of a sample that is fully representative of the target population over multiple waves of a panel study. Moreover, there may be reductions in other costs of data collection that partially if not completely balance the extra incentive costs.

If the incentive had been increased to \$50 for all sample members in all four strata, and if the response rate in each stratum increased to that obtained for those who were in the \$50 treatment group, approximately 918 additional interviews would have been obtained, and the overall non-response rate would have decreased from 12.1% down to 8.2%, or almost a third. The additional cost of the higher incentives would have been about \$654,000, or about \$712 per additional interview. The cost and the impact of the higher incentive varied by stratum. For every expected additional interview in stratum 1, the cost would have been \$931; for stratum 2, \$961; for stratum 3, \$372; and for stratum 4, \$771. Thus, the greatest cost-benefit ratio would likely have been achieved by offering the higher incentive to households in which there was non-response at the previous wave.

To look at the possible savings in other field costs that could be achieved by giving higher incentives to all respondents, the number of face-to-face and telephone contacts were examined separately for coupled and uncoupled households in each of the four strata and in each of the three incentive groups. The expected number of each type of contact in each of those cells were then calculated as if the average number of calls had been that observed for those given \$50 for that type of household. A preliminary estimate is that the number of telephone contacts would have been reduced by 8.7%, and the number of face-to-face calls by 14.4%. The implication is that the number of interviewer hours would have been substantially reduced if \$50 had been given to all respondents, and that the extra cost of the higher incentives would have been partly offset by the reduction in interviewer pay, travel expenses, and other charges that are linked to interviewer hours.

Discussion

This paper has looked at only a limited 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 response rate increases with the size of the incentive, and also that the number of contacts that

interviewers make before finalizing a household decreases as the size of the incentive increases. The implication is that panel attrition can be reduced by increasing the incentive offered to the sample members. Experience in future waves of HRS will provide evidence about the longer-term consequences of this experimental variation. 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 implication, 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. On the other hand, increasing the incentive may raise the expectations of sample members, and this could lead to lower participation at later waves if those expectations are not met. Another danger is that higher incentives could change the perception of some respondents about the meaning of the incentive, from a "token of appreciation" for their fulfillment of what they perceive as a civic obligation to payment for their time, in which case the incentive could be converted to an hourly rate and compared to what they earn with the rest of their time. These questions can be addressed after future waves of HRS have been completed.

Another concern is with respect to the quality of the data, including the number of questions that are not answered 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. This concern could, and should, be addressed by analysis of data from the 2000 wave of HRS for those in the different incentive treatment groups.

A final observation is that the focus of this paper has been on a single wave of a panel study and the effects of increased incentives on the response rate at that wave. More important for the quality of data from a panel study is the cumulative attrition and the extent to which that can be reduced by design features such as increased respondent incentives. If we make the assumption that the response rates for previous wave respondents and non-respondents would be maintained at the levels observed in the 2000 wave of HRS for those in the \$20 treatment group, the response rate after five waves would decline by about two per-

centage points (from about 88.5% to 86.6%), whereas if the incentive were increased to \$50 for all sample members and the response rates for previous wave respondents and non-respondents were maintained at the levels observed in the 2000 wave of HRS for those in the \$50 treatment group, the response rate after five waves would increase by about six percentage points (from about 88.5% to 93.7%). Of course, the incentive level pales in importance compared to another aspect of the HRS design, namely that non-respondents at one wave are not automatically dropped from the subsequent wave. If the response rate for previous wave respondents were maintained at the level observed for those in the \$50 treatment group but previous wave non-respondents were dropped, the response rate after five waves would decline by more than twelve percentage points (from about 88.5% down to 76.1%).

References

- Heberlein, T.A., and R Baumgarten. 1978. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review* 3:447-62.
- Juster, F.T. and R.M. Suzman. 1995. "An Overview of the Health and Retirement Study." *Journal of Human Resources* 30 (Suppl.):S7-S56.
- Singer, E. 2002. "The Use of Incentives to Reduce Nonresponse in Household Surveys." In *Survey Nonresponse*, ed. R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, pp. 179-93. New York: Wiley.
- Singer, E., J. Van Hoewyk, N. Gebler, T. Raghunathan, and K. McGonagle. 1999. "The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys." *Journal of Official Statistics* 15:217-30.
- Soldo, B.J., M.D. Hurd, W.L. Rodgers, and R.B. Wallace. 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.
- Yu, J., and H. Cooper. 1983. "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires." *Journal of Marketing Research* 20:36-44.

Table 1: Response rates in 2000, by Incentive Groups and by strata

Stratum	Incentive Provided			Total
	\$20	\$30	\$50	
1) Deteriorating health	93.4% 1,296	94.6% (ns) 299	96.7% (*ns) 302	94.1% (ns) 1,897
2) Proxy informant in 1998	91.7% 2,096	89.1% (ns) 304	94.4% (ns/*) 304	91.8% (ns) 2,704
3) Non-interview in 1998	37.4% 2,148	38.0% (ns) 300	45.2% (**/ns) 294	38.3% (*) 2,742
4) Everyone else	94.6% 15,242	97.5% (**) 606	98.5% (***/ns) 609	94.8% (***) 16,457
TOTAL	88.3% 20,782	83.4% 1,509	86.9% 1,509	87.9% 23,800

Significance levels are indicated in parentheses. In the \$30 column, the significance is with respect to the difference in the response rate for those given \$30 vs. those given \$20. In the \$50 column, the first significance (before the slash) is with respect to the difference in the response rate for those given \$50 vs. those given \$20, while the second is with respect to the difference in the response rate for those given \$50 vs. those given \$30. In the Total column, the significance is for the overall differences across the three groups. The significance levels are as follows:

- ns: $p > .05$
- *: $p < .05$
- ** : $p < .01$
- ***: $p < .001$

Table 2: Logistic Regressions to Response vs. Non-response, for Respondents Who Were and Were Not Part of a Couple in 2000

	Uncoupled	Couples: At Least One Interview	Couples with At Least One Interview: Two Interviews
Sample size	7,652	8,074	7,238
Incentive level (vs. \$20)			
\$30 vs. \$20	ns	ns	**
\$30	-.029 (.164)	.063 (.158)	.726 (.279)
\$50 vs. \$30	*	*	ns
\$50	.415 (.173)	.499 (.172)	.589 (.260)
\$50 vs. \$20	*	**	*
Stratum (vs. stratum 4)			
1. Deteriorating health	-.064 (.181)	-.167 (.211)	-.628 (.251)
2. Proxy interview	-.518 (.171)	-.570 (.144)	-.753 (.191)
3. Non-interview	-3.245 (.095)	-3.479 (.099)	-3.208 (.147)
Constant	2.724 (.055)	3.233 (.073)	3.875 (.101)

Significance levels indicated for tests of differences between incentive levels:

- ns: $p > .05$
- *: $p < .05$
- ** : $p < .01$
- ***: $p < .001$