1. INTRODUCTION AND SUMMARY

This paper presents findings from an experiment designed to test the effectiveness of providing monetary payments to survey respondents in order to improve their responses to personal interviews. A $5 payment per interview was offered to a randomly selected portion of a survey panel of economically disadvantaged youths for three waves of interviews undertaken in conjunction with an evaluation of the Job Corps program. The effects of these respondent payments were then tested for improving the quantity of responses (search efficacy and interview completions) and the quality of responses (interview completions and item nonresponse).

The beneficial effects of monetary payments to respondents in surveys with personal interviews have been hypothesized to stem primarily from the increased cooperation of sample members and of individuals who can be of assistance in locating them. Also, respondent payments are expected to lead to increased productivity of interviewers who feel they are making less of an imposition on respondents when respondents are paid. However, little is known about the magnitude of the effects of respondent payments in surveys with personal interviews.

Monetary incentives have been shown to be effective in increasing the quantity of responses from mailings and follow-up surveys whose burdens are substantially greater than a personal interview, such as those which ask the respondent to submit to a physical examination (see Miller, Kennedy, and Bryant 1972). Similarly, monetary incentives have been shown to be effective in getting respondents to accept greater burden in personal interviews, such as agreeing to complete more modules in return for an additional payment (see Chromy and Horvitz 1978). However, previous empirical research has been inconclusive about the effects of payments on search efficacy and interview completions for personal interviews, due largely to small sample sizes and the consequent imprecision of estimates (see Perber and Sudman 1974).

Furthermore, for data quality, the current state of research leaves even the direction of the effect--positive or negative--in serious doubt (see Cannell and Henson 1974; Perber and Sudman 1974; Friedman and Augustine 1979; and Hansen 1980). While it is hypothesized that monetary payments will help respondents provide more thoughtful and accurate answers to questions, sample members who are motivated to respond because of the payment may provide responses that are of particularly low quality.

The evaluation of the economic impact of Job Corps provided a large national sample for which monetary incentives were expected to be effective. This sample included a longitudinal panel of over 5,800 economically disadvantaged youths, approximately 70 percent of whom were offered payments on the basis of random assignments in an experimental design. The effects of monetary incentives to respondents are hypothesized to be larger in magnitude for repeated waves of interviews and for samples of populations that are difficult to locate and interview--both of which apply to this experiment. Thus, a large number of experimental observations are available for a national population for whom monetary incentives are expected to be particularly effective.

As expected, the impacts of monetary incentives on the efficacy of the search effort and interview completions increased over the successive waves of interviews. By the third wave of interviews, the observed impacts on four measures of search-effort success and interview completions were statistically significant and important in magnitude. For every 100 sample members who were to be paid, we received 10 more responses to an advance mailing than for a similar number of sample members who were not paid, which suggests some clear effects from respondent payments in terms of search efficiency and completion rates with mail surveys. Furthermore, even though the overall completion rates for the second follow-up were very high, we obtained approximately 5 additional completions per 100 sample members in the payment group.

The estimated impacts of monetary incentives on item nonresponse were most pronounced at baseline interview, and declined over time. By the second follow-up survey, they were very small in magnitude and were statistically insignificant. This seemingly anomalous pattern of effects can be attributed at least in part to the questions becoming easier to answer over time and to the least cooperative respondents dropping out of the nonpayment group over time.

The payment experiment shows some clear benefits to monetary incentives. However, precise inferences cannot be drawn about the size of these effects beyond our specific experiment with $5 payments to disadvantaged youths and using an intensive search process. The size of effects will undoubtedly vary by the amount of the payment, by the characteristics of the underlying population, and by the intensity of search procedures.

The next two sections describe in more detail the experimental setting and design, respectively. The fourth section outlines the field procedures that were followed. The experimental findings are reported in Sections 5 and 6--first for search efficacy and interview completions, and then for the quality of the data. Finally, in Section 7, we summarize our conclusions from the experiment.

2. THE SETTING FOR THE EXPERIMENT

The Job Corps program provides job training, basic education, and related services in a residential setting to economically disadvantaged youths between the ages of 16 and 21. Our design for the evaluation of the economic impact of Job Corps included a personal baseline interview that was administered in early 1977 to Corpsmembers...
who were then enrolled in centers and a series of personal follow-up interviews administered periodically after these Corpsmembers left the program. The baseline interviews obtained information on Corpsmembers' previous work experience, training, education, and related behavior, as well as on their demographic and socioeconomic backgrounds. The follow-up interviews extended these data with information on the postprogram experiences of Corpsmembers, so that at each follow-up we attempted to reinterview those Corpsmembers who had been out of Job Corps for a sufficient period of time. In addition to Corpsmembers, we interviewed on the same schedule a comparable group of disadvantaged youths who had not attempted to enroll in Job Corps.

Observations were obtained from two area probability samples—one for Corpsmembers and another for the comparison group. The Job Corps sample included all Corpsmembers who were in the program during May 1977 and who lived in three-digit zip-code areas (the Primary Sampling Units, PSUs) that were randomly selected. The comparison group was drawn in 15 geographic sites scattered throughout the country. These comparison PSUs were also selected based on zip-code areas but with a stratified probability procedure designed to yield a random sample of sites similar to the areas from which Corpsmembers came and to oversample locations in which Job Corps did not recruit heavily in 1977. Within comparison sites, youths were randomly selected from lists of school dropouts and employment service registrants, stratified to yield a probability sample similar on average to Corpsmembers in terms of age, educational level, length of time out of school, race-ethnicity, poverty status, and related socioeconomic characteristics.

3. THE EXPERIMENTAL DESIGN

The objective of the payments experiment was to assess whether or not monetary incentives could be an effective survey tool for increasing the quantity and quality of responses to personal interviews. The design of this experiment was constrained in three ways: (1) the underlying population was limited to economically disadvantaged youths, (2) the payments experiment was a secondary objective of the overall project, and (3) only a small amount of resources could be used for fielding and analyzing the payments experiment. It was within the context of the above objective and constraints that the experimental-design issues were decided.

Amount and Type of Payment. Because of limitations with the underlying population and the available sample, we decided to limit the payments to a fixed amount of money. Varying the amount or type of payment to learn more about monetary incentives would have been very useful. However, we first wanted to be able to learn whether monetary payments could be effective, and the responses from our limited population would not allow accurate estimates for more general populations. Also, the implicit sample sizes were too small to support experimentation with several alternative payment amounts and mechanisms. The amount and type of payment per interview were set at $5 checks or money orders, based on expectations that this would lead to measurable effects on survey response for a longitudinal sample of economically disadvantaged youths and a personal interview of moderate length (approximately one-half hour).

Sample Size and Allocation. The design decisions on the appropriate sample size and allocation also reflect the constraints under which the payments experiment was conducted. First, the overall sample size was determined by the primary objective of the study—i.e., evaluating the economic impact of Job Corps. Second, in allocating this sample to payment and nonpayment statuses, the needs of the payments experiment had to be traded against the needs of the primary objective of the study. On the one hand, the experiment with payments would have been served best by assigning 50 percent of the sample to payment and 50 percent to nonpayment status. On the other hand, the preferred option for meeting the primary objective of the evaluation would have been to assign the entire sample to payment status.

Our resolution of this sample allocation tradeoff was to assign 70 percent of the sample to payment and 30 percent to nonpayment status. With this sample allocation, adequate statistical power was obtained to detect payment effects that were large enough in magnitude to be relevant without jeopardizing the primary objective of the evaluation. For the follow-up surveys, if the "true" effect for a proportionate variable were 0.050, for example, we would have a reasonably good chance of observing a statistically significant finding (approximately 90 percent power for a 10 percent significance-level test), even though the overall sample would be a relatively small 0.015 (i.e., 0.30 times 0.050 equals 0.015).

Experimental Assignments. A random procedure was used to assign sample to either payment or nonpayment status based on the geographic areas which formed the PSUs. This area clustering of payments was adopted for three reasons. First, some sample members lived close together within geographic areas. Thus, varying experimental assignments within sites would likely have caused biases from information being spread to some youths in the nonpayment group that others were receiving payments. Second, an area saturation of payments was desired in most areas to engender community support when attempting to locate respondents. Third, an area probability assignment was convenient, because both the Job Corps and comparison groups had been selected in that way at baseline and the same PSUs could be used.

4. FIELD PROCEDURES FOR THE EXPERIMENT

The surveys underlying this experiment consisted of a baseline and two follow-up interviews with the baseline respondents forming the sampling frame for subsequent follow-ups. At baseline we attempted to interview a probability sample of Corpsmembers who were then in residence at centers and a probability sample of comparison youths. In each of the two follow-up surveys we attempted to reinterview everyone in the comparison group who had been interviewed at baseline and those Corpsmembers interviewed at baseline who had been out of Job Corps for a
sufficient period of time.
The baseline questionnaire was designed to collect detailed information on seven topics: socioeconomic background, Corpsmembers' evaluations of Job Corps, demographic characteristics, employment and income, education and training, criminal behavior, and addresses for survey tracking. The first follow-up obtained updated information on the last six of these topics and the second follow-up obtained updated information on the last five.

All of the baseline interviews were administered in person--for Corpsmembers, at their centers; for comparison youths, at their homes. No prior contact or notification of the baseline survey was given to sample members, so those who were to receive payments were informed of them through the search process or when contacted.6/ However, a more serious limitation on the payments experiment at baseline was that only comparison-group members could be paid at that time, because the Corpsmember sample was still residing at centers and receiving pay allowances from Job Corps. It was expected that payments to Corpsmembers at baseline would be unnecessary and hence ineffective while they were in residence at centers (substantiated by a 97% completion rate). More importantly, it was feared that such payments would be disruptive and results would be biased by the presence of both payment and nonpayment samples at the same centers. If, instead, the experimental assignments were made on the basis of centers, the intermingling of payment and nonpayment samples would have been substantial for the follow-up surveys.

The relatively small sample sizes for the payments experiment at baseline, as shown in Table 1, reflect the fact that the payments experiment was instituted only in comparison sites at that time. The sample sizes for search efficacy and interview completions are based on the number of attempted interviews, and those for data quality are based on full interviews. Furthermore, only 9% of the comparison sample was rural, and search efficacy was expected to (and did) differ substantially between rural and urban sites with greater variation in the rural sites. All five rural sites were assigned to payment status because they would not contribute much information on the effectiveness of payments and this ensured better data for the evaluation. Consequently, youths in three urban comparison sites were randomly assigned to nonpayment status at baseline with geographically stratified assignments used to increase the efficiency of statistical comparisons.

The analysis of search efficacy and interview completions at baseline contrasts the experience in the three nonpayment sites with that in five urban sites where monetary incentives were offered. Two urban payment sites were omitted from this analysis because of special fielding complications, which could bias estimates (one site because of prior notification and the other because of erroneous addresses). The samples for analysis of data quality include youths in these latter two urban sites and the five rural sites, because the respective fielding and geographic differences should have little effect once an interview is begun.

The first follow-up survey was conducted approximately nine months after baseline and included everyone in the comparison group who completed the baseline interview and everyone in the Corpsmember group who both completed the baseline interview and had been out of Job Corps for at least four months. All the interviews for the first follow-up were again administered in person and primarily in the youths' homes. However, there were two major changes that affected the payments experiment at the first follow-up--(1) prior notification of the potential monetary payment was given to everyone who was to be paid and (2) the Corpsmember group was included in the experiment.

In an advance mailing to all potential respondents alerting them to the first follow-up and requesting address updates, we notified those in the payment group that they would be paid for interviews. In addition, the comparison sample members in the payment group were told at baseline that they would again receive $5 payments if and when they were reinterviewed.

The second follow-up survey was conducted approximately 14 months after the first and made use of most of the same procedures. However, there were two notable changes--(1) additional Corpsmembers from baseline had left Job Corps and were included in the second follow-up and (2) the interviewing was changed from in-person to a mixed-mode procedure of telephone attempts followed by in-person attempts for a portion (over 60%) of those who could not be interviewed by telephone.

Both the inclusion of additional Corpsmembers and the mixed-mode interviewing for the second follow-up complicate the analysis of the payments experiment. The inclusion of additional Corpsmembers means that even by the time of the second follow-up some sample members were being paid for interviews for the first time. The mixed-mode interviewing means that the results for the second follow-up cannot be compared directly to those for the first follow-up and baseline. Consequently, we simplify interpretations, we have limited the second follow-up analysis to observations of sample members who were interviewed at both earlier surveys and who were to be attempted in person if the telephone attempts were unsuccessful.6/

5. FINDINGS ON QUANTITY OF INTERVIEWS

Table 2 summarizes findings on search efficacy and interview completions based on differences in sample means between the payment and nonpayment groups. In addition to estimates of differences in group means, for the follow-up surveys the availability of baseline data enabled us to compute regression estimates that control for other differences between the two samples and to test for differences in effects among subgroups. However, the regression estimates of overall effects are nearly identical to the sample mean estimates reported here, as expected with a classic experimental design. Also, we did not find any differences in effects across subgroups with the regression estimates.

Search Efficacy. Two variables can be used to measure search efficacy--(1) the proportion of the sample that returned postcards from the advance letters and (2) the proportion of the
A. For Analysis of Search Efficacy and Interview Completions

B. For Analysis of Data Quality

Sample payment sample. However, by the second follow-up, there was a large and statistically significant increase in the proportion of the payment sample that returned these postcards, as compared to the nonpayment sample. However, by the second follow-up, there was a large and statistically significant increase for the payment group. For the first follow-up (see Table 2) there was only a very small and marginally significant increase in the proportion of the payment sample that returned these postcards, as compared to the nonpayment sample. However, by the second follow-up, there was a large and statistically significant increase for the payment group. For the second follow-up, the sample mean difference between payment and nonpayment groups shows an increase of just over 1 returned postcard for every 100 sample members, from a base of approximately 27 returned postcards per 100 sample members in the nonpayment group. For the second follow-up, the difference is 1 returned postcard for every 100 sample members, with a statistical significance of one percent. Furthermore, this estimate for the second follow-up is associated with a small sample base of only 17 postcards, as compared to the nonpayment group. These findings suggest that, in a longitudinal study, monetary payments can reduce the costs of locating respondents and increase the responses to mail surveys.

The second measure of search efficacy is the percent of the sample that was located. For the baseline survey, we were able to locate just over 71 percent of the nonpayment sample—a low base because of the inaccurate addresses originally obtained for the comparison sample. We found a relatively large (6 percentage point) and statistically significant increase in the proportion of the sample located for the payment group (up from 0.714 to 0.774). However, the statistical significance is biased upward because of the extreme clustering and the consequent potential for systematic interviewer error with the baseline findings (the clustering effect is substantially less for the follow-up data). For the first follow-up survey, we were able to relocate approximately 87 percent of the sample and found virtually no difference in our ability to locate youths in either the payment group or the nonpayment group. By the time of the second follow-up, however, our ability to locate youths in the nonpayment group began to decline substantially, as opposed to maintaining nearly an 87 percent rate of success in locating the youths in the payment group. This difference in our ability to locate youths for the second follow-up was approximately 4 percentage points (87 percent as opposed to 83 percent) and statistically significant.

In summary, both of our measures of search efficacy show some clear gains to respondent payments for a longitudinal survey. However, the gains for a single cross-sectional survey are less.

Interview Completions. Once a sample member is located, the only reason for not completing an interview is a refusal on the part of the respondent. In this section we consider empirical evidence on the effects of respondent payments on refusal rates and on the proportion of the sample with usable data.

The overall refusal rates were very low for our survey (always less than 3%) because we made a great effort to convert refusals. Therefore, the payment effect on refusal rates could not be very large in magnitude. For the baseline and first follow-up surveys, there was virtually no difference in the refusal rate between the payment and nonpayment groups. For the second follow-up survey, there was nearly a 2 percentage point lower refusal rate for the payment group, which was statistically significant and reduced the refusals by nearly half. The refusal rate increased over time for the nonpayment group but held relatively constant for the payment group.

The estimated payment effect on refusal rates was very high, we obtained approximately 5 additional completions per 100 sample members in the payment group.

6. FINDINGS ON DATA QUALITY

We do not have data from secondary sources with which to directly validate survey data. However, we do have two types of proxy measures for data quality: (1) interview completion rates and (2) item nonresponse in completed interviews. A higher quantity of responses, as found above for the payment group in the second follow-up, should lead to higher quality data through reduced nonresponse bias. Measures of item nonresponse ("Don't know," refused, and blank answers) are expected to reflect the amount of thought and accuracy respondents give to difficult questions.

As shown in Table 3, the estimated impacts of payments on item nonresponse were most pronounced at baseline and declined over time. By the
TABLE 2
SEARCH EFFICACY AND INTERVIEW COMPLETIONS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline Non-Payment Group</th>
<th>First Follow-Up Non-Payment Group</th>
<th>Second Follow-Up Non-Payment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of sample that returned postcards</td>
<td>n.a.</td>
<td>.277</td>
<td>.271</td>
</tr>
<tr>
<td>Proportion of sample located</td>
<td>.774</td>
<td>.960***</td>
<td>.875</td>
</tr>
<tr>
<td>Proportion located who refused the interview</td>
<td>.029</td>
<td>.017</td>
<td>.023</td>
</tr>
<tr>
<td>Proportion of sample with usable data</td>
<td>.751</td>
<td>.696</td>
<td>.847</td>
</tr>
</tbody>
</table>

TABLE 3
DATA QUALITY

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline Non-Payment Group</th>
<th>First Follow-Up Non-Payment Group</th>
<th>Second Follow-Up Non-Payment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Don't know&quot; answers per completed interview</td>
<td>3.119</td>
<td>.608</td>
<td>.469</td>
</tr>
<tr>
<td>Refused answers per completed interview</td>
<td>.331</td>
<td>.068</td>
<td>.232</td>
</tr>
<tr>
<td>&quot;Don't know&quot; or refused answers per complete</td>
<td>3.450</td>
<td>.676</td>
<td>.701</td>
</tr>
<tr>
<td>Blank answers per completed interview</td>
<td>.815</td>
<td>1.602</td>
<td>.772</td>
</tr>
<tr>
<td>Total Item nonresponse per complete</td>
<td>4.265</td>
<td>3.279</td>
<td>.703</td>
</tr>
</tbody>
</table>

Notes to Tables 2 and 3:
*Significantly different from zero at the 20% level of significance (10% for a one-tail test).
**Significantly different from zero at the 10% level of significance (5% for a one-tail test).
***Significantly different from zero at the 5% level of significance (2.5% for a one-tail test).
****Significantly different from zero at the 1% level of significance (0.5% for a one-tail test).
n.a. means not applicable.

second follow-up, they were very small in magnitude and statistically insignificant. This seemingly anomalous pattern of effects can be attributed at least in part to the questions becoming easier to answer over time (less complicated recall, more concrete questions, and increased familiarity on the part of interviewers and respondents) and to the least cooperative respondents dropping out of the nonpayment group over time.

The results for the number of "Don't know" answers per completed interview are the most striking and conclusive. The estimated effect of payments at baseline is in the expected direction, large, and statistically significant. The magnitude of the effect diminishes for the first follow-up and virtually disappears by the second. At baseline, the effect is to reduce the number of "Don't know" responses by three for every four completed interviews. The average number of such nonresponses at baseline is nearly four per interview in the nonpayment group. By the first follow-up, the difference fell to just over one "Don't know" response in every three interviews, but the base had also fallen dramatically to under one in 10 interviews (as the questions became easier to answer). The base falls even further for the second follow-up, and the difference is close to zero and positive.

The estimated payment effect on refusals to answer individual questions is similar to those for "Don't know" answers. However, the patterns for refused answers are not as pronounced, nor are any of the individual differences highly significant.

Differences in total item nonresponse at baseline were partially offset by counter-intuitive but statistically significant differences in the opposite direction for blank answers, which are a combination of interviewer errors and uninterpretable answers. It is difficult to see how payments would increase the number of blank answers, especially because it can be shown that they were caused primarily by interviewers neglecting to indicate in the main questionnaire that they had written answers to confidential questions on a separate form.

Total item nonresponse shows a payment-related difference at baseline of just under one fewer nonresponse per completed interview for the payment group. The base is slightly over five nonresponses per interview for the nonpayment group. At the first follow-up, the difference is only slightly smaller and is statistically significant, even though the base is just over half of what it was at baseline. At the second follow-up, the
base falls to approximately two item nonresponses in every three interviews, and the experimental difference disappears.

What emerges from our findings is some tentative evidence that monetary incentives improve data quality. The primary measures on which this conclusion is based are the proportion of completed interviews and the number of "Don't know" responses per completed interview. The number of completed interviews was significantly higher for the payment group at the third interview, and evidence from the first two waves of interviews indicates that there were beneficial effects of payments on item nonresponses.

7. CONCLUSIONS

The findings from our experiment with payments to survey respondents go far in establishing the effectiveness of such payments. The hypotheses that payments will improve search efficacy and interview completions for longitudinal surveys are verified here for our sample of disadvantaged youths. After one or two interviews, our ability to locate potential respondents and to obtain data from those who were located deteriorated in the absence of monetary incentives, but not when $5 payments were offered to respondents. In addition, payment effects on the willingness of sample members to return postcards from advance letters suggest that monetary incentives may be an effective aid for mail surveys and for reducing the cost of locating respondents for personal interviews. The estimated impacts on item nonresponse were most pronounced at baseline and declined over time.

The payment experiment shows some clear benefits to monetary incentives. However, precise inferences cannot be drawn about the size of effects beyond our specific experiment with $5 payments to disadvantaged youths and using an intensive search process. The size of effects will undoubtedly vary by amount of payment, characteristics of the population, and intensity of search procedures.

FOOTNOTES

1 Funding for the research reported in this paper was provided by the Office of Program Evaluation of the Employment and Training Administration of the U.S. Department of Labor.

2 For more details on the evaluation, see Mallar et al. (1980).

3 Youths in the age range of Corpsmembers (16 to 21) and with their economically disadvantaged backgrounds are generally very mobile and difficult to locate. From Census data, approximately 50% of them can be expected to move in any given year, with 50% of these moves being outside their initial metropolitan area. In a previous, one-time survey of former Corpsmembers (see Harris 1969), interviewers were able to locate only 60% of the sample, and full interviews were obtained only from 48% of the sample.

4 Sample members who did not respond at baseline were not followed. All those who responded at baseline were followed in subsequent waves even if they did not respond to an intervening interview.

5 For more details on the sample designs for the Job Corps and comparison groups and the statistical methodology used to draw inferences about program impacts, see Mallar et al. (1980).

6 In the process of searching for comparison youths in the payments sample, information was made readily available on the monetary benefit to being interviewed.

7 Estimates were also computed for the full sample at second follow-up and for the telephone attempts. No substantial deviations were found compared to the results presented in this paper.

REFERENCES


