Journal of Official Statistics, Vol. 11, No. 4, 1995, pp. 413–431

Computer-assisted Personal Interviewing: An Experimental Evaluation of Data Quality and Cost

Reginald P. Baker, Norman M. Bradburn, and Robert A. Johnson[1]

This article uses experimental data from the National Longitudinal Survey (NLS) to compare the quality and costs of data collected using computer-assisted personal interviewing (CAPI) and traditional paper-and-pencil personal interviewing (PAPI). The experiment was carried out as part of the NLS Youth Cohort (NLS/Y), a continuing longitudinal face-to-face survey of United States residents who were aged 14-21 in 1979. Interviewers were randomly assigned to conduct personal interviews using the same NLS/Y questionnaire as adapted for administration by either PAPI or CAPI. The results suggest that, compared with PAPI, CAPI yields lower item nonresponse rates and greater respondent willingness to disclose sensitive information. Few distributions of responses obtained using CAPI differ substantially from those obtained by administering the same questionnaire using PAPI. The respondent burden, as measured by the length of an interview, is about 20% less with CAPI. CAPI may initially be somewhat more expensive than PAPI, but the cost difference is likely to narrow as organizations and interviewers gain experience in using CAPI, especially if the costs of portable computers continue to decline.

Key words: Technical feasibility; portable computers; stratified sampling; random assignment; matching; field management; completion rate; skip pattern; missing data; mode effects; response effects.

1. Introduction

Computer-assisted personal interviewing (CAPI) extends to face-to-face interviewing the technologies developed and deployed for computer-assisted telephone interviewing (CATI) in the late 1970s and early 1980s. CAPI questionnaires are designed and implemented in software systems with all of the functionality now common to CATI. (See for example, Nicholls and Groves 1986; Martin, O'Muircheartaigh, and Curtice 1993). The software then is loaded on to a small, light, portable computer so that interviews can be conducted almost anywhere–in respondents' homes, in places of business, in public places such as hospitals, libraries, and restaurants, and even outdoors on porches, in backyards, and public parks.

Proponents of CAPI believe that it offers a number of important advantages over conventional paper-and-pencil interviewing (PAPI), most notably higher overall data

[1] National Opinion Research Center, University of Chicago, 1155 E. 60th Street, Chicago, IL 60637, U.S.A.

Acknowledgements: The authors wish to thank Martin R. Frankel, Michael R. Pergamit, Alisu Schoua-Glusberg, Randy Olsen and his staff, Edwin Hunt and Mark Wojick for their contributions to the design and execution of this experient. Funding for the study and its analysis came from the U.S. Bureau of Labor Statistics, the Russell Sage Foundation and NORC.

quality and quicker turnaround of interview data at a lower cost. (For discussion, see Baker 1990; Martin et al. 1993.) Although CATI technology has been widely used for over a decade, CAPI has been slower to emerge because of five basic concerns: technical feasibility, respondent acceptance, interviewer acceptance, effects on data quality, and costs. Steady improvements in both software and hardware and several years of international testing and research have put at least the first three of these to rest (Bradburn et al., 1991b; NCHS 1988; Lyberg 1985; Keller, Metz, and Bethlehem 1990; Foxon 1987; Bernard 1989; Martin et al. 1993; O'Reilly, Hubbard, Lessler, Biemer, and Turner 1994). There are now several commercially-available software systems which can meet most CAPI requirements. Portable computers weighing seven pounds and less are now on the market and have proven that they can withstand the demands of field interviewing with CAPI.

CAPI has recently been adopted as the standard personal interviewing technique in the Current Population Survey (CPS), a monthly survey conducted by the U.S. Bureau of the Census for the U.S. Bureau of Labor Statistics (BLS) that is the primary source of information on labor force characteristics of the U.S. population. The U.S. Census Bureau's expectation is that the implementation of CAPI in the CPS will "improve data quality and reduce respondent burden" (Copeland and Rothgeb 1990). However, the Census Bureau's evaluations were not designed to separate the effects of CAPI from the effects of a new CPS questionnaire that was implemented at the same time as the switch to CAPI (January 1994) (Rothgeb and Creighton 1991). CAPI has also been recently adopted in other continuing government surveys, such as the labour force and family resources surveys in Britain (Martin et al. 1993), even though there remain many unanswered questions about the data quality and costs of CAPI.

This article uses data from a controlled experiment conducted as part of the National Longitudinal Survey (NLS) to examine differences between CAPI and PAPI in data quality and costs. Section 2 describes the sample design and the controlled experiment. Sections 3–5 compare CAPI and PAPI with respect to different aspects of data quality. Section 3 provides evidence that CAPI yields lower item nonresponse rates. Section 4 provides some evidence that CAPI respondents are more forthright than PAPI respondents in responding to sensitive questions. Section 5 summarizes other significant model differences in response distributions. Section 6 compares costs under CAPI and PAPI. Section 7 summarizes our conclusions.

2. Sample and Experimental Design

The National Longitudinal Survey of Labor Market Experience/Youth Cohort (NLS/Y) is a longitudinal face-to-face survey of the United States population now in its 17th year. The survey is funded by the U.S. Bureau of Labor Statistics (BLS). The Center for Human Resources Research at Ohio State University (CHRR) and the National Opinion Research Center at the University of Chicago (NORC) conduct the survey in cooperation with BLS. The sample consists of 11,464 people who were aged 14-21 as of January 1, 1979 and who have been interviewed every year since. The sample was stratified by sex, race, ethnicity and poverty status with oversampling of

Blacks, Hispanics, white youths in poverty and equal numbers of men and women in each group. The questionnaire is primarily oriented toward labor market participation, education and fertility. It is an extremely complex questionnaire with many skip patterns and extensive rostering. The complexity of the questionnaire and the difficulties that interviewers have in making their way through it make it an ideal candidate for a computer-assisted administration.

Randomized experiments comparing CAPI and PAPI were embedded in each of the 1989 (Round 11) and 1990 (Round 12) rounds of NLS/Y. In the Round 11 experiment, half of the Ohio sample for the 1989 round (Round 11), randomly selected from all Ohio cases, was interviewed using CAPI. In all, 301 cases were completed by CAPI (completion rate of 91.8%) and 264 cases were completed by PAPI (completion rate of 95.6%). The success of the Round 11 pilot led BLS to approve a larger experiment using CAPI for the full 1990 round (Round 12).

The Round 12 experiment to compare CAPI and PAPI was designed with two constraints. First, it was to be a true experiment and strict procedures were developed to ensure that there was no contamination between the experimental and control groups. Second, the experiment should not compromise the quality of the data for the study as a whole, including the overall completion rate target of 92%, a completion rate which was necessary to maintain the integrity of the survey panel.

The experimental design called for division of the NLS/Y sample into three subsamples:

- Twenty-five percent of the respondents were to be interviewed using CAPI.
- An additional 25% were designated as control cases to be interviewed using PAPI.
- The remaining 50% were regular NLS/Y cases, interviewed using PAPI but not part of the experiment.

Data from the CAPI and PAPI control cases will be compared in the analyses of mode effects. The design serves as a substitute for overlapping samples which one might use in a repeated cross-sectional survey. Thus, any mode effects of CAPI which cannot be corrected can be controlled for in any future use of the data.

Because simple random selection of cases into the two groups would not result in cost efficient clustering of cases for interviewer assignment, mode selection was based on the attributes of the interviewers and their assignment characteristics, for example, geographical location and previous NLS/Y experience. Twenty-five percent (77) of the Round 12 interviewers were randomly assigned to the CAPI condition, 25% (77) were assigned to the control condition and the remaining 170 interviewers were assigned to the remaining cases which were also conducted using PAPI.

CHRR developed the interviewing software for the survey. NORC developed the management system, including the telecommunications component that interviewers used to transmit their cases to the NORC central office (Speizer and Dougherty 1991). NORC interviewers used Compaq LTE notebook computers, each equipped with a 20 MB hard disk and weighing about seven pounds.

Field managers assigned NLS/Y cases to interviewers as in any normal round. Interviewers were characterized by assignment size and complexity, urban versus

rural case load, assignment type (regular versus regular-plus-conversion cases), and ethnicity (Black, white, or Hispanic). Interviewers with Spanish language cases and cases outside the 48 contiguous states were excluded from the experiment for cost reasons, the anticipated difficulty with international phone transmission of cases, and the cost of providing interviewer support by international long distance. These types of cases are a very small proportion of the NLS/Y sample and excluding them from the experiment was judged to have little adverse effect on the generalizability of the experimental findings.

Interviewers were randomly assigned to conditions by using procedures which balanced the conditions on the basis of interviewer type: regular interviewer versus converter; ethnicity; geographic region; and metropolitan/non-metropolitan areas. This was accomplished by sorting the interviewers with respect to these four variables and then forming groups of four "matched" interviewers. Within each matched group, interviewers were randomly distributed among four groups: one CAPI group, one Control group, and two regular groups.

Cases belonging to a CAPI condition interviewer were done in the CAPI mode, those belonging to a control interviewer were in the Control condition which used PAPI. The remaining cases were "other" and used PAPI. To maintain experimental integrity interviewers and cases could not cross mode; that is, a CAPI interviewer could not interview a respondent in any other condition as long as the interviewer was participating in the experiment. No control interviewers were allowed to do cases using CAPI.

The total number of cases assigned to the CAPI mode was 2814. The Control sample had 2715 cases (See Table 1).

NORC's automatic Field Management System (FMS) is designed to track, on a case by case basis, production, cost and sample information. We modified the system for Round 12 to allow tracking of the experiment so that no contamination occurred between experimental conditions.

Near the end of the field period it became clear that, within the constraints of the budget, we could not achieve the targeted completion rate for the entire study and still maintain strict adherence to the experimental conditions. Thus, for cost reasons, we stopped the experiment after we had reached a completion rate of at least 80% each for the CAPI (81.9%) and the Control (84.2%) conditions. Interviewers completed the remaining cases by whatever method was convenient.

3. Missing Data

A central feature of computer-assisted interviewing is that it can help prevent

Table 1. The NLS/Y Round 12 experimental design

	PAPI	CAPI
Interviewers	77	77
Assigned cases	2715	2814
Completions	2287	2306
Completion rate	84.2%	81.9%

interviewers from making mistakes. The most obvious example is following the skip pattern of the questionnaire. Properly programmed CAPI software ensures that the interviewer always follows the correct skip pattern; items cannot be skipped or left blank.

The research record seems to bear this out. For example, in a small-scale CAPI feasibility study conducted by Research Triangle Institute in 1987 found that over 90% of the errors made by paper-and-pencil interviewers were failures to record an answer; CAPI interviewers made no such errors (Sebestik, Zelon, DeWitt, O'Reilly, and McGowan 1988). Groves and Mathiowetz (1984) demonstrated that CATI interviewers were able to follow complex skip logic virtually flawlessly, while paper-and-pencil interviewers made a great many mistakes, almost five times as many as with CATI. Though based on a small sample, the results of O'Reilly et al. (1994) suggest similar results may apply to computer-assisted self-administration.

Olsen (1991), using data from all rounds of NLS, including the CAPI experimental data from Rounds 11 and 12, looked at the three kinds of missing data—incorrect skips, refusals, and don't know responses. Table 2 shows the average rate of incorrect skips, refusals, and don't know for the first ten rounds of the NLS/Y. It compares these means with similarly computed means for Rounds 11 and 12 but using only the data collected with CAPI. Similar to findings reported by Martin et al. (1994), the results suggest CAPI has little or no effect on refusals and don't know responses. The most important finding is that CAPI completely eliminated illegal skips, which account for most of the missing data in the NLS/Y.

4. Sensitive Questions

One of the most interesting mode differences among face-to-face, self-administered and telephone interviewing found by Sudman and Bradburn (1974) in their review of response effects was on sensitive questions. They found that in more private modes of administration, such as self-administration, respondents were more likely to report negative behavior. While it is not apparent that CAPI would be viewed as more anonymous that PAPI, since both are face-to-face, it is possible that because the interviewer is entering the responses into a computer rather than writing them down on a form that has the respondent's identifying information on it, respondents feel more anonymous, which encourages them to report more of what they might consider to be negative or sensitive information.

One of the most oft-cited research findings in the CAPI literature is Waterton and Duffy's (1984) study of Scottish drinking habits in which it was reported that

Table 2. Mean percent of items coded as refused, don't know, and illegal skip in NLS/Y

	Rounds 1-10	CAPI
Refused	.053%	.045%
Don't know	.368%	.380%
Illegal skip	.994%	–

Base: 39,677,478(rnds1–10) 8,623,603(CAPI)
Source: Olsen(1991)

self-administered computer-assisted interviews yielded higher estimates of alcohol consumption, a behavior which is traditionally underreported. Our analysis of the NLS/Y Round 11 feasibility test data showed a similar effect. Twenty-three of 26 questions about alcohol-related problems had higher frequencies for CAPI than for PAPI, although only two were statistically significant (Bradburn et al. 1991a). One hypothesis for these findings is that respondents perceive computer-assisted interviews as more confidential and therefore are more likely to report truthfully. The results of O'Reilly et al. (1994) on the reporting of illegal drug use in self-administered interviews also suggested that computer-assisted interviews encourage more complete reporting.

Unfortunately, the Round 12 NLS/Y questionnaire did not carry over the questions on drinking from Round 11. However, there are a number of questions on use of contraceptives which we examined to see whether there were differences between the two modes. In one series of questions males were asked first whether they or their partner have used any birth control methods in the last month. Those responding yes were asked how often. Respondents who said they had used a method were asked which of 15 different methods they used. These same questions were repeated for females.

Table 3 compares PAPI and CAPI responses to the two birth control questions separately for males and females. Sixty-six percent of the male CAPI respondents report having used some form of birth control in the previous month compared to 58.5% for controls. When males reporting having used a method in the last month were asked how often they used it, 89.5% of the CAPI respondents reported that they always used it compared to 93% for the controls. While the differences in both cases are small, they are statistically significant. Seventy-two percent of the female CAPI respondents report having used birth control in the last month compared to 68% of the controls. This difference is statistically significant. Unlike males, females show no reporting differences across mode when asked how often.

In addition to the questions presented in Table 3, respondents were also asked about specific birth control methods, and females were also asked a large number of questions on current and former pregnancies, including questions about

Table 3. Responses to questions on birth control

		MALES		FEMALES	
QUESTION	RESPONSE	PAPI%	CAPI%	PAPI%	CAPI%
During the last month have you or your partner/spouse used any form of birth control? ($n = 4067$)	Yes (versus No)	58.5*	66.1*	68.3*	72.4*
In the past month, how often have you or your partner/spouse used birth control? ($n = 2683$)	Always (versus Less often)	93.0*	89.5*	95.0*	94.7*

* CAPI vs. PAPI difference is significant at .05 level based on two-sample t-test.

health-related behavior such as smoking, drinking, and drug use during pregnancy. Our analysis has not uncovered any mode differences in these items, although, especially for questions involving former pregnancies, the skip pattern results in very few cases for analysis.

The hypothesis that respondents perceive computer-assisted interviewing as more confidential and therefore are more likely to report sensitive information is consistent with the results of a self-administered interview of CAPI respondents immediately after the main interview. Respondents were asked to compare the experience of being interviewed by computer with the experience of the previous eleven rounds of interviews by paper-and-pencil and to indicate which they felt was the more confidential. Forty-seven percent felt that the CAPI interview was more confidential with only 5% favoring paper-and-pencil. The remaining 48% felt that the two modes were equally confidential (Wojcik and Baker 1992).

5. Other Modal Differences in Response Distributions.

In addition to the sensitive questions discussed in the preceding section, we compared CAPI and PAPI distributions of responses to a total of 445 NLS/Y questionnaire items covering the following nine topic areas: marital status and history, education, military background, labor force, employment, job training, health, income, and immigration. The 445 questionnaire items gave rise to 385 categorical response variables, covering the nine topic areas, and 60 continuous response variables, primarily dealing with income and assets but also including other items related to work and employment conditions, e.g., number of days of paid vacation and number of employees at the respondent's place of work. Of the 385 categorical variables, 48 (12.5%) yielded statistically significant CAPI-PAPI differences at the .05 level based on chi-square tests of independence. Of the 60 continuous variables, 12 (20%) yielded statistically significant CAPI-PAPI differences at the .05 level of significance based on two-sample t-tests adjusted for unequal variances between groups. Thus the overall percent of significant CAPI-PAPI differences equals about 13.5%.

Each significant chi-square result was based on a contingency table satisfying a minimum criterion for convergence, i.e., at least 50% of the cells had fitted frequencies greater than 5. Even after judicious collapsing of response categories, we found that about 20% of the 385 contingency tables failed to satisfy this criterion because the questions were asked only to a small subclass of respondents. Each t-test was based on a sample of at least 80 in each mode. Prior to testing for a CAPI-PAPI difference in means, we trimmed extreme values from each of the CAPI and PAPI subsamples according to a uniform criterion that depended on the particular test variable.

Using significance tests as a screen for substantively important findings when examining a large number of variables will produce a number of chance findings. In particular, even if there were no real CAPI vs. PAPI differences in response distributions, we would expect about 5% of the items tested to manifest statistically significant modal differences at the .05 level of significance. Thus observed differences in this experiment must be treated cautiously, and some may not replicate in another

experiment. Moreover, the test statistics are adjusted neither for complex sampling nor for the clustering of CAPI and PAPI cases within interviewer assignments. This implies the true significance level of most tests is likely to be greater than .05.

Table 4 summarizes our results for the 19 categorical variables that gave rise to CAPI-PAPI differences that were statistically significant at the .05 level and, in our judgement, substantively significant as well. These 19 differences also serve to summarize the main patterns discovered in looking over the results for all 48 categorical variables that gave rise to significant chi-squares. Of these 48 categorical variables with significant CAPI vs. PAPI differences, 13 involved questions in which the PAPI field interviewers had to refer back to a previous question to check for an answer and on the basis of this manual check make a decision about which further questions to ask. CAPI eliminated this error prone checking by automatically skipping to the appropriate question. In Table 4, variables that required the PAPI field interviewer (FI) to refer to a previous question are described as "FI coded."

For each significant variable, Table 4 shows the set of eligible respondents, the response category (ordinarily the category that gave rise to the largest modal difference), the PAPI and CAPI percents responding in the category, and, in parenthesis, the "Base n", i.e., the number of eligible CAPI or PAPI respondents who answered the question. Due to skip patterns and item nonresponse, the Base n's for each question are smaller than the corresponding numbers of complete cases. As shown in Table 1, the number of complete cases equals 2306 for PAPI and 2287 for CAPI.

The categorical variables with significant CAPI-PAPI differences in Table 4 are grouped in five sets:

1. Household roster questions: CAPI cases had a small, but significantly higher percentage of spouses listed in the household roster (CAPI = 54%; PAPI = 51%). This difference would appear to be due to the more careful updating of the household roster that is done in CAPI. In the PAPI mode, the FI has a printed face sheet listing the members of the household from the last interview. The FI goes through a series of questions to ask for new members or departures from the household. The FI has the entire list of possible new or departing member categories in front of her and is supposed to ask each question, but could skip some questions if the respondent says things that suggest that questions to come have already been answered. There is only one copy of the face sheet which is corrected manually as is appropriate.

In CAPI, each question about a household member appears on a separate screen and must be paged through one at a time. At the end of each iteration of additions, departures, and correction of age and education status, a new household roster is displayed and the FI is directed to check its accuracy with the respondent. It seems likely that this complete presentation of the roster three times with an opportunity to check it with the respondent would produce fewer omissions or miscoding of a household member than the manual version.

Although the difference is not substantively very large, it seems likely that it is the more unusual household arrangements that produce this discrepancy. In subsequent panels of Table 4, there are some differences between CAPI and PAPI responses on questions and spouse's behavior that might be due to selection bias for questions that are only asked when there is a spouse listed on the roster.

2. *Labor force questions*: Similar to the findings of Bergman, Kristiansson, Olofsson, and Säfström (1994) in comparing CATI and PAPI responses to labor force questions, we found PAPI and CAPI response distributions were remarkably similar for most such questions. An apparently important modal difference occurs in response to a question on respondent's activity last week. This NLS/Y question corresponds exactly to Q19 of the pre-1994 CPS questionnaire. While there is no difference between modes in the proportion reporting that they were working, there is a difference in some of the remaining categories which yields an overall statistically significant difference between modes. The largest differences are for the category: "With a job, but not at work" (PAPI = 5%; CAPI = 3%) and "Other" (PAPI = 4% and CAPI = 6%).

The biggest difference in treatment between modes is that there is central office editing of the questionnaire in PAPI, but not in CAPI. Thus an interview coded "other" could be edited and the code changed to a smaller code on the basis of either the verbatim response or subsequent questions which follow-up for those able to work, but not currently working. The "other" code in CAPI is not edited and is transmitted directly to CHRR.

That this editing is probably what accounts for the difference is seen when one traces out the pattern of responses after the initial response to CPS Q19 for those eligible to work (codes 2, 3, 4, 5, and 7). These follow-up questions exactly correspond to two CPS questions: CPS Q20, asking whether the respondent did any work at all last week, and CPS Q21, asking the FI whether the respondent was coded "with a job but not at work" on Q19, and, if not, asking the respondent whether he or she was temporarily absent or laid off last week. If one adds those respondents who say they were temporarily absent or laid off last week to those who said in Q19 that they were "with a job, but not at work," then the modal difference disappears (PAPI = 6.5%; CAPI = 6.3%).

3. *Earnings questions*: Although there has been considerable research on interviewer reactions to CAPI, it has mostly been directed at their willingness and ability to adapt to new interviewing technology. There has been relatively little attention paid to how interviewers change their behavior in the actual interview situation. Some proponents of CAPI emphisize that the technology will bring greater control of interviewers and thereby eliminate undesirable features of interviewer behavior, while others argue its major benefit is to give interviewers an improved tool for doing what they always have done.

Our analysis has uncovered one potential change in how interviewers present questions to respondents. The questionnaire has a sequence of questions aimed at establishing wage rates for respondents in 1989. In a multi-part question respondents are first asked how much they earn at their current job. Then they are asked the unit of pay (per hour, per week, and so forth) for the just reported wage. Those who report a unit of pay other than hourly are routed to a question which asks whether they are paid by the hour, and if so, how much they earn per hour. All of these questions are on a single page of the questionnaire. A similar series of questions is used to establish the wage rate of the respondent's spouse in 1989.

Both the questions on earnings of respondent and those on earnings of spouse

showed significant differences between CAPI and PAPI in the unit of pay that was initially chosen by the respondent to report earnings. In both series of questions, PAPI picked up more initial reports of earnings per hour, while CAPI picked up more initial reports of earnings per year. However, in the subsequent question asked of everyone who did not report earnings (respondent's or spouse's) on a per hour basis, the CAPI-PAPI difference was reversed with the result that combining the proportions reporting per hour in the two questions substantially reduces the CAPI-PAPI difference. That is, combining the two question yields closely similar PAPI and CAPI estimates of the proportion paid by the hour both in the case of respondent's earnings (PAPI = 64%, CAPI = 64%) and in the case of spouse's earnings (PAPI = 56%, CAPI = 51%). Only the latter difference is statistically significant.

Once all reports of earnings are converted to a per hour basis, neither respondent's nor spouse's estimated wage rate differs significantly by mode. For example, the estimated wage rate equals $8.67 per hour for spouses of CAPI respondents and $8.64 per hour for spouses of PAPI respondents.

It is not clear why PAPI respondents were more likely to report wages on an hourly basis, but the same finding emerged in the Round 11 NLS/Y experimental pretest, so it may be taken as a real difference. The earnings question did appear differently in the two modes. In PAPI, the three questions appeared on one page and the interviewer may have discerned that the ultimate purpose was to find out who was paid by the hour and the hourly wage rate. In CAPI, the questions came on the screen one at a time, so the interviewer could not have known there would be subsequent questions to determine whether or not the spouse was paid by the hour. Thus, it seems likely that some PAPI interviewers anticipated the following questions and obtained and entered all the hourly information at once, allowing them to skip over the following question.

4. Sources of health plans: Four questions on sources of health plans showed small but significant differences between PAPI and CAPI. These questions are about the sources of health care coverage of the respondent, the respondent's spouse, or the respondent's children. All of the differences are around 1 percentage point with CAPI obtaining the higher percent in all cases. There are no obvious format differences between PAPI and CAPI on these questions.

5. Questions about employment: Three questions regarding employment benefits, asked of those who worked last week, exhibited mode effects. Two of them (life insurance and maternity leave) were reported more frequently by PAPI respondents (PAPI = 53%; CAPI = 50% for life insurance; PAPI = 48%; CAPI = 46% for maternity leave), while more CAPI respondents reported subsidized parking as a benefit (PAPI = 48%; CAPI = 55%). There is no obvious explanation for these differences. In the CAPI version, each benefit appears on a separate screen rather than as a list on one page as in PAPI. Also, in CAPI the question runs further along the screen in fewer lines perhaps making it more likely that the entire question for life insurance and maternity benefits was read.

Four other employment questions showed significant mode differences. Two involved questions about the respondent's supervisor (How the supervisor got the position and whether the supervisor decided the respondent's pay) and two involved

the respondent's job (Whether earnings are based on performance and whether respondent is responsible for deciding other people's pay or promotion). There is nothing about the layout of these questions that suggests any hypothesis to account for the differences.

The final row of Table 4 shows that PAPI cases recorded fewer instances in which there were "gaps" in the respondent's employment history, i.e., periods of non-employment, than did CAPI cases (PAPI = 30%, CAPI = 37%). For this question in both modes, the interviewer had to draw lines manually on an employment calendar and then count up the number of lines. In this manual operation it would be easy to miss a "gap" and not draw a line. In CAPI but not in PAPI, the automatic fill-in directs the interviewer to a question about a gap, where it has been recorded earlier in the interview, and automatically fills in the dates of the gap. Thus, if the CAPI interviewer has forgotten to draw a line to indicate a gap, the computer automatically reminds her of the gap and of the date of the gap. There is no such reminder in PAPI. Correctly answering the question on employment gaps also required a correct interpretation of gaps in the calendar, and there were many more missing data for the question about employment gaps in PAPI than in CAPI. Out of 259 cases of missing data to the question about gaps, 225 are in PAPI.

Table 5 presents results for 4 of the 12 continuous variables that evidenced significant mean differences between CAPI and PAPI based on two-sample t-tests. These 4 were the largest and most substantively interesting of the mean differences we uncovered:

1. Respondent's income in 1989: On average, CAPI respondents reported higher incomes in 1989 than PAPI respondents (PAPI = \$15,154, CAPI = \$15,950). This difference is on the borderline of statistical significance at the .05 level. There is no statistically significant modal difference when respondents who reported no income in 1989 are dropped from the analysis. A somewhat larger proportion of PAPI respondents than of CAPI respondents reported no income (PAPI = 17.5%, CAPI = 16.4%). Some respondents might report no income to avoid having to give a correct answer to this sensitive question. If so, this result provides further evidence that CAPI respondents are more willing to disclose sensitive information than PAPI respondents (Section 4).

2. Dollars per week spent on food used at home: Respondents who did not receive food stamps were asked how much they spent on food used at home in an average week. They could respond either for a week or for a month. Most respondents responded for a week. After trimming outlying values (>\$500 per week), we found no significant CAPI-PAPI difference among respondents who responded for a week (PAPI = \$59.8, CAPI = \$59.4). For the minority of respondents who responded for a month, we found the average PAPI weekly expenditure to be significantly higher than the average CAPI weekly expenditure (PAPI = \$44.7, CAPI = \$36.4). Once the monthly data are transformed into weekly amounts and added to the amounts initially reported by week, there is no significant modal difference (PAPI = \$57.4, CAPI = \$57.5).

3. Dollars spent eating out: CAPI respondents reported more dollars spent eating out than PAPI respondents (PAPI = \$19 per week, CAPI = \$23.6 per week). The

*Table 4. Categorical variables with significant CAPI-PAPI differences**

VARIABLE	ELIGIBLE TO RESPOND	RESPONSE CATEGORY	PAPI % (Base *n*)	CAPI % (Base *n*)
1. HOUSEHOLD ROSTER QUESTIONS:				
R's spouse on HH roster? (FI coded)	All	Yes	50.7% (2270)	54.1% (2280)
2. LABOR FORCE QUESTIONS:				
R's activity last week (CPS Q19)	All	1. Working	69.9%*	70.1%*
		2. With job not at work	5.1%	3.0%
		3. Looking for work	2.7%*	1.9%*
		4. Keeping house	15.0%*	15.9%*
		5. Going to school	1.5%*	1.5%*
		6. Unable to work	1.5%*	1.2%*
		7. Other	4.3%	6.5%
			(2254)	(2271)
Any work at all last week? (CPS Q20)	R's activity last wk. coded 2, 3, 4, 5, or 7.	No	89.1% (644)	86.1% (653)
R "with job not at work"? (CPS Q21)	R's who reported no work last week	No	81.4% (574)	90.2% (562)
Temporarily absent or laid off last week? (CPS Q21)	No work last wk.; Not "with a job not at work"	Yes	6.9% (467)	14.8% (507)
3. EARNINGS QUESTIONS:				
R's choice of unit of pay	R worked	per hour	46.7% (1980)	38.0% (2003)
Spouse's choice of unit of pay	R married in 1989; Spouse worked	per hour	28.1% (1053)	15.4% (1077)
4. SOURCES OF HEALTH PLANS (Code-all-that-apply questions)				
Source of R's health plan	R covered by health plan	Spouse's previous employer	0.3% (1808)	1.1% (1861)
Source of spouse's plan	Spouse covered by plan	R's previous employer	0.5% (1002)	1.3% (1074)
Source of child's plan	Child covered by plan	R's previous employer	0.4% (1074)	1.4% (1072)
Source of child's plan	Child covered by plan	Spouse's previous employer	0.7% (1074)	1.6% (1072)
5. QUESTIONS ABOUT EMPLOYMENT:				
Employer provides life ensurance?	R worked last wk.	Yes	68.2% (1728)	63.7% (1794)

Table 4. Continued

VARIABLE	ELIGIBLE TO RESPOND	RESPONSE CATEGORY	PAPI % (Base *n*)	CAPI % (Base *n*)
5. QUESTIONS ABOUT EMPLOYMENT (continued)				
Employer provides maternity leave?	R worked last week	Yes	62.8% (1584)	58.6% (1734)
Employer provides or subsidizes parking?	R worked last wk.	Yes	61.6% (1766)	70.4% (1808)
Earnings based on performance?	R worked last wk.	Yes	19.9% (1986)	25.5% (2003).
R responsible for others' pay or promotion?	R worked 30 or more hrs., supervises others	Yes- Fully	6.8% (601)	12.2% (649)
How did R's supervisor get to his/her position?	R worked 30 or more hrs. last wk.	Worked way up through the ranks	54.7% (1562)	51.5% (1647)
R's supervisor responsible for deciding R's pay?	R worked 30 or more hrs. last wk.	Yes- Fully	30.4% (1669)	34.9% (1686)
Gaps in employment - yes/no (FI coded)	All	Yes	30.0% (2032)	36.5% (2302)

* With the exception of those differences marked using asterisks (see text), each difference between CAPI and PAPI percentages reported in this table is statistically significant at the .05 level based on a likelihood-ratio chi-square test of independence with one degree of freedom.

difference is based on the combined responses of respondents who reported by the week or those who reported by the month. In this case, there is a modal difference when the zero's and extreme values are treated as missing, but again when the totals for week and month are combined and when zero's and extreme values and treated as missing, there is no significant difference in the total amount reported (PAPI = \$20.2, CAPI = \$21.1).

4. Cash assets: There was a modal difference in reported cash or liquid assets. The difference was marginally significant for the totals and became significant when extreme values (>\$3,000,000) were omitted from the analysis (PAPI = \$6,871, CAPI = \$11,081). Even omitting all values greater than \$30,000 from the analysis did not eliminate this difference or make it statistically insignificant (PAPI = \$3,799, CAPI = \$4,300). It is possible that this is the case of CAPI being viewed as more confidential and obtaining more accurate reports of cash assets.

Despite the apparent CAPI-PAPI differences shown in Tables 4 and 5, what is

especially remarkable about the overall pattern of results, given the fairly large sample sizes and liberal criteria of statistical significance, is the large number of critical NLS/Y variables that yielded small or statistically insignificant CAPI-PAPI differences. These insignificant variables, not shown in Tables 4 and 5, included standard questionnaire variables for measuring marital status, changes in marital status, living arrangements, history of living arrangements, hours worked, educational attainment, current school enrollment, reasons for leaving school, current and previous military service, hours worked in the week prior to the survey (CPS Q20A), main reason for working less than 35 hours (CPS Q20C), overtime and extra hours (CPS Q20E), periods of being laid off (CPS Q21A, Q21B), looking for work (CPS Q22), employer characteristics (CPS Q23), employment goals (CPS Q24), usual hours of work (CPS Q25), prior and current job training, work-related disabilities, sources of income, immigration status, and many others.

In addition to mean differences, we examined the differences in variances. Of the 48 comparisons, 26 showed significant differences in variances, but there was no observable pattern related to mode. For 12 variables, PAPI produced higher variances; for 14 variables, CAPI had higher variances. Thus, there does not appear to be any consistent modal effect on the variances.

In summary, most significant modal differences occurred in questions where the FI in PAPI cases had to do some manual checking of previous questions, face sheets or the calendar, in order to know what subsequent questions to ask. In complex questionnaires such as the NLS/Y, such manual procedures are error prone. The automatic checking and fill-ins that occur in CAPI are designed to reduce this type of error.

*Table 5. Continuous variables with significant CAPI-PAPI differences**

VARIABLE	ELIGIBLE TO RESPOND	UNIT OF MEASURE: RANGE	PAPI mean (Base *n*)	CAPI mean (Base *n*)
R's income in 1989	All	$ (0,99999)	$15,154 (2240)	$15,950 (2247)
$ spent per week on food used at home (R and spouse)	R's who responded in $ per week	$ (0,500) per week	$59.8/wk.* (1567)	$59.4/wk.* (1878)
	R's who responded in $ per month	$ (0,500) per week	$44.7/wk. (209)	$36.4/wk. (87)
$ per week spent eating out (R and spouse)	R's who responded in $ per week or $ per mo. (combined)	$ (0,500) per week	$19.0/wk. (2218)	$23.6/wk. (1947)
Cash assets (R and spouse)	R reported some assets	$ (1,3000000)	$6,871 (1534)	$11,081 (1475)

* With the exception of those differences marked using asterisks (see text), each difference between CAPI and PAPI means reported in this table is statistically significant at the .05 level based on a two-sample *t*-test of difference-in-means. The degrees of freedom for this test equals $(n_1 + n_2 - 2)$, where n_1 and n_2 are the base *n*'s of CAPI and PAPI eleigible respondents. Given the large sample sizes, shown in parentheses in the last two columns, the test statistics were compared to the standard normal distribution rather than to the *t* distribution. (See, e.g., Snedecor and W. Cochran (1980) pp. 93, 469).

6. The Cost of CAPI

The reduction of survey costs has often been mentioned as a likely benefit of CAPI. In theory, CAPI completely eliminates the post-interview data processing required by paper-and-pencil interviewing, including the substantial costs of keying, machine-editing, and computer programming to set up and execute these tasks. Yet skeptics have warned that any savings in post-interview data processing due to CAPI might be more than offset by increased interviewing costs and by increased pre-interview costs, including the costs of portable computers for use in the field, CAPI software design, and training interviewers in the use of the CAPI software. Unfortunately, research on CAPI has seldom focused on costs, preferring to consider overall feasibility, respondent acceptance, and interviewer acceptance.

An exception is the RTI field test reported by Sebestik et al. (1988). These authors report higher costs of fielding a survey with CAPI than with PAPI, even though hardware and software costs were excluded, largely because hoped-for savings in data processing did not materialize while the costs of interviewing rose substantially. The NLS/Y experience was somewhat different. NLS/Y training costs were about 70% higher with CAPI, compared with an 18% increase reported by Sebestik et al. (1988). NLS/Y direct field interviewing costs were about the same as with CAPI and PAPI, compared with a 17% increase reported by Sebastik et al. (1988). The main difference between NLS/Y and the field test reported by Sebestik et al. (1988) is that, in NLS/Y, overall data processing costs were more than 2.5 times higher for PAPI than for CAPI. Overall, when the amortized costs of the laptop computers are added to the field costs, the total direct cost per completed case in NLS/Y was about 6% higher with CAPI than with PAPI (Table 6).

There were other programmer support costs that were required to field the experiment which should be taken into account, but we find it difficult to know how many of them to attribute to the experimental nature of this study and how many would be attributable to CAPI on an ongoing basis. When a reasonable allocation of these costs are added, our best estimate is that the total direct costs of CAPI cases in the NLS/Y is about 12% higher than the PAPI cases. As more interviewers have experience with CAPI, we expect that training costs will go down. If hardware costs also continue to decline, the differential between PAPI and CAPI modes should diminish.

The slightly higher CAPI field interviewing costs in NLS/Y were due to an added

Table 6. Cost comparison by mode (approximate direct costs/interview)

	Mode	
Task	PAPI	CAPI
Training	6.95	11.77
Field data collection	118.61	119.89
Data preparation	44.19	11.99
Subtotal	169.75	143.65
Laptop computers	–	36.73
Total	169.75	180.88
$n =$	2287	2306

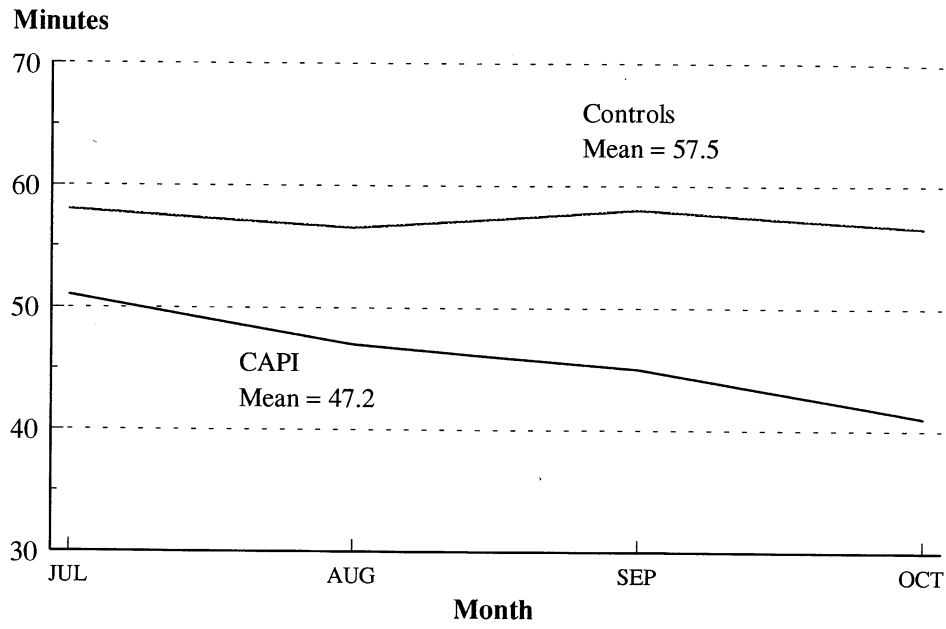

Fig. 1. Mean length of interview by month

support burden the transition to CAPI requires, rather than to increased interviewing time. Indeed, the average length of an interview in the Round 12 NLS/Y was much shorter with CAPI than with PAPI.

Figure 1 plots the average length of the interviews conducted each month of the Round 12 field period by mode. The timings for the CAPI interviews are based on two time stamps generated by the software. The starting time is logged at the first question after the household enumeration has been completed and verified; the ending time is generated after the last question, before the locating and interviewer remark sections. In the PAPI cases, interviewers directly recorded the time in those places in the questionnaire.

In the first month of interviewing, the timings for CAPI were about six minutes less than those for the Controls. Over the entire field period, as interviewers became more adept at CAPI the average length of the CAPI interviews declined to a point where by the last month of interviewing the CAPI interviewers were on average almost 15 minutes shorter than those conducted with paper-and-pencil. Over the entire field period, the Control interviews lasted about 57 minutes while the CAPI interviews averaged only 47 minutes.

Should we interpret this finding as indicative of the potential for CAPI to save time (and therefore interviewing costs) as well as to reduce respondent burden? Or does the finding rather point to a pacing problem in CAPI, where the interview moves too quickly, controlled by a machine rather than the human pace of conversation?

We favor the former interpretation, because the complex NLS/Y questionnaire requires the PAPI interviewer to do much more than simply read the questions and record the answers. For example, there are more than 60 different instructions to the interviewer to go back to a complex facesheet, find a specific bit of information,

record it in the questionnaire, and then use this information in the question. There are nearly 30 occasions when the interviewer must refer to an answer given prior to the preceding question in order to determine the next question to be asked. There are nine questions which require that the interviewer perform an arithmetic calculation. All these operations are done automatically, instantly, and correctly by the CAPI software.

We urge caution in generalizing the NLS/Y experience to other surveys. The NLS/Y incurs large data processing costs prior to the field work because of the extensive generation of assignment lists, complex facesheets, and locating information, and large data processing costs after the field work because of extremely rigorous post-data-entry batch cleaning. The complex NLS/Y questionnaire perhaps makes it ideal for administration using CAPI. In experiments using a less complicated questionnaire, Martin et al. (1993) found CAPI interviewers took longer on average than PAPI interviews, with the difference plausibly attributable to lack of interviewer experience with CAPI. More generally, we expect that several key cost elements—need for interviewer training, interviewer field support, and the cost of hardware—will decline as organizations gain experience and hardware costs continue to fall. In the end, the final cost differences between modes will depend on the mix of elements in a given survey.

7. Conclusions.

We conclude from our experiments that computer-assisted personal interviewing (CAPI) offers many advantages in improved data quality over traditional paper-and-pencil (PAPI) methods. Its main advantage is in reducing item nonresponse rates due to interviewer branching errors in questionnaires that have complicated skip patterns. This advantage will be particularly important in studies involving extremely complex questionnaires where there is multiple branching contingent on the respondents' answers to previous questions.

In addition, there is some evidence that respondents view CAPI as providing more assurances of confidentiality than PAPI and thus are more forthright in responding to sensitive questions. Unfortunately the particular round of the NLS/Y in which the experiments were conducted did not contain enough questions of a sensitive nature to establish this conclusion fully. The evidence, however, is sufficiently compelling to encourage additional research to test the hypothesis of improved candor on sensitive questions.

The advantages of CAPI over PAPI appear to be obtained with small mode effects on fairly simple, non-sensitive questions. Relatively few response distributions varied substantially by mode of administration. Most modal differences could be attributed to a reduction in the information-retrieval demands placed on the interviewer using CAPI and suggest that the CAPI responses may have had greater validity.

While surveys using CAPI may initially be somewhat more expensive than comparable surveys using PAPI, the cost difference is likely to decline as interviewers gain experience in using CAPI. This is especially true for large surveys featuring questionnaires with complex skip patterns, since such surveys are likely to take full advantage

of the technical capabilities of CAPI. For studies in which there are large data capture costs with PAPI, the cost savings from the immediate data entry capabilities of CAPI may offset most, if not all of the increased costs of programming the CAPI questionnaire, interviewer training, and the laptops.

One word of caution. The advantages of CAPI in eliminating interviewer errors in executing branching instructions assume that adequate testing of the CAPI programmed questionnaire has been conducted. Obviously, if a branching error is programmed into the questionnaire, item non-response for those items may well be greater than in the case of PAPI. Extensive testing of the representation of the questionnaire in the computer is essential to take advantage of CAPI. Careless testing can result in greater, not fewer errors.

8. References

Baker, R.P. (1990). What We Know About CAPI: Its Advantages and Disadvantages. Paper presented at the 1990 American Association of Public Opinion Research Conference, Lancaster, PA, 17-20 May.

Bergman, L., Kristiansson, K., Olofsson, A., and Säfström, M. (1994). Decentralized CATI Versus Paper and Pencil Interviewing: Effects on the Results in the Swedish Labor Force Surveys. Journal of Official Statistics, 10, 181–196.

Bernard, C. (1989). Survey Data Collection Using Laptop Computers. INSEE Report No 01/C520. Paris.

Bardburn, N.M., Frankel, M.R., Baker, R.P., and Pergamit, M.R. (1991a). A Comparison of Computer-Assisted Personal Interviews (CAPI) with Paper-and-Pencil (PAPI) Interviews on the National Longitudinal Study of Youth. Paper presented at the 1991 American Association for Public Opinion Research Conference, Phoenix, AZ, 17-19 May.

Bradburn, N.M., Frankel, M.R., Hunt, E., Ingels, J., Schoua-Glusberg, A., Wojcik, M., and Pergamit, M.R. (1991b). A Comparison of Computer-Assisted Personal Interviews with Personal Interviews in the National Longitudinal Survey of Labor Market Behavior–Youth Cohort. Proceedings of the U.S. Bureau of the Census Annual Research Conference, 389–397.

Copeland, K. and Rothgeb, J.M. (1990). Testing Alternative Questionnaires for the Current Population Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association, 63–71.

Dippo, C.S., Tucker, C., and Valliant, R. (1993). Survey Methods Research at the U.S. Bureau of Labor Statistics. Journal of Official Statistics, 9, 121–135.

Foxon, J. (1987). Field Trials in the Labour Force Survey Lap-top Computer Project in November 1987. Survey Methodology Bulletin, July, 12–25.

Groves, R.M. and Mathiowetz, N.A. (1984). Computer-Assisted Telephone Interviewing: Effects on Interviewers and Respondents. Public Opinion Quarterly, 48, 356–369.

Ingels, J., Mikros, P., and Cooke, M. (1989). CAPI at NORC: Working Paper No. 1. Chicago, IL: NORC.

Keller, W., Metz, K.J., and Bethlehem, J.G. (1990). The Impact of Microcomputers

on Survey Processing at the Netherlands Central Bureau of Statistics. Proceedings of the U.S. Bureau of the Census Annual Research Conference, 637–675.

Lyberg, L. (1985). Plans for Computer Assisted Data Collection at Statistics Sweden. Bulletin of the International Statistical Institute, Book III, Topic 18.2.

Martin, J., O'Muircheartaigh, C., and Curtice, J. (1993). The Use of CAPI for Attitude Surveys: An Experimental Comparison with Traditional Methods. Journal of Official Statistics, 9, 641–662.

National Center for Health Statistics (NCHS) (1988). Report of the 1987 Automated National Health Interview Survey Feasibility Study. Working Paper No. 32, Hyattsville, MD.

Nicholls, W.L. and Groves, R.M. (1986). The Status of Computer-Assisted Telephone Interviewing: Part I-Introduction and Impact on Cost and Timeliness of Survey Data. Journal of Official Statistics, 2, 93-115.

Olsen, R.J. (1991). The Effects of Computer Assisted Interviewing on Data Quality. Unpublished report, Columbus, OH: CHRR.

O'Reilly, J., Hubbard, M., Lessler, J., Biemer, R., and Turner, C. (1994). Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection. Journal of Official Statistics, 10, 197–214.

Rothgeb, J.M. and Creighton, K.P. (1991). Development of the Proposed Revised Current Population Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association, 56–65.

Sebestik, J.H., Zelon, H., DeWitt, D., O'Reilly, J.R., and McGowan, K. (1988). Initial Experiences with CAPI. Proceedings of the Forth Annual Research Conference, U.S. Bureau of the Census, 357–365.

Snedecor, G. and Cochran, W. (1980). Statistical Methods, 7th. ed. Ames, IA: Iowa State University.

Speizer, H. and Dougherty, D. (1991). Automating Data Transmission and Case Management Functions for a Nationwide CAPI Study. Proceedings of the U.S. Bureau of the Census Annual Research Conference, 375–388.

Sudman, S. and Bradburn, N.M. (1974). Response Effects in Surveys: A Review and Synthesis. Chicago, IL: Aldine Publishing Co.

Waterton, J.J. and Duffy, J.C. (1984). A Comparison of Computer Interviewing Techniques and Traditional Methods in the Collection of Self-Report Alcohol Consumption Data in a Field Study. International Statistical Review, 52, 173–182.

Wojcik, M. and Baker, R.P. (1992). Interviewer and Respondent Acceptance of CAPI. Proceedings of the U.S. Bureau of the Census Annual Research Conference, 613–621.

Received March 1994
Revised October 1995