

## CAPI: AN EXPERIMENTAL EVALUATION

Reg Baker, Norman Bradburn, and Robert Johnson, National Opinion Research Center  
Robert A. Johnson, NORC, 1350 Conn. Ave., NW, Suite 500, Washington, D.C. 20036

**Key words:** Skip pattern; missing data; mode effects; response effects

CAPI (computer-assisted personal interviewing) extends to face-to-face interviewing the CATI (computer-assisted telephone interviewing) technologies developed and deployed 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). The software then is loaded onto a small, portable computer so interviews can be conducted almost anywhere--in respondent's homes, places of business, public places such as hospitals, libraries, and restaurants, and even outdoors on porches, in backyards, and public parks.

CAPI is widely believed to offer at least three advantages relative to PAPI (paper-and-pencil personal interviewing): CAPI makes the data available sooner, lowers data collection costs, and results in better quality data. This paper uses experimental data from the National Longitudinal Survey Youth Cohort (NLS/Y) to compare the data quality and costs of CAPI and PAPI. We first present the survey and experimental designs. Second, we present experimental comparisons of CAPI and PAPI on four measures of data quality: interview length, missing data rates, response differences on sensitive questions, and response differences on other questions. Third, we present a comparison of CAPI and PAPI survey costs.

### Survey and Experimental Design

The National Longitudinal Survey of Labor Market Experience/Youth Cohort (NLS/Y) is a longitudinal face-to-face survey now in its 16th year. The survey is funded by the Bureau of Labor Statistics (BLS). The Center for Human Resources Research (CHRR) at Ohio State University is the prime contractor and NORC the data collection subcontractor. The sample consists of 11,464 people who were aged 14-21 as of Jan. 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 a complex questionnaire with many skip patterns and extensive roster-

ing. 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. The main experiment was conducted in the 1990 round (Round 12). In Round 12, 77 of 324 interviewers in the U.S. (about 25%) were randomly assigned to conduct CAPI interviews, and another 77 randomly assigned to conduct PAPI interviews. (The remaining 170 interviewers also conducted PAPI interviews but were not part of the experiment.)

The random assignment of interviewers to modes was stratified so that the two treatment groups were balanced with respect to geographic location, experience, and ethnicity. Interview completion rates of greater than 80% were obtained in both treatment groups. As a result, the comparisons of this paper are based on data for approximately 2300 completed interviews in each treatment group.

An important feature of the experimental design is that interviewers and not sample cases were the experimental units. Precision and statistical power would have been greater if we had randomly assigned sample cases to modes within interviewer assignments, but this approach would also have reduced the verisimilitude of the treatments. Regardless of mode, interviewers learn skills by repeating the same procedures. This process would have been harmed if interviewers had been obliged to frequently change their mode of interviewing.

### Length of Interview

The average length of interviews equalled about 57 minutes for interviews conducted using PAPI and about 47 minutes for interviews using CAPI. The Round 12 field period lasted from July through October 1990. As interviewers became more experienced in using CAPI, we found the average length of CAPI interviews declined from about 51 minutes to about 41 minutes. By the end of the field period, CAPI interviews were averaging about 15 minutes shorter than PAPI interviews.

We think the reductions in interview length were associated with improvements in data quality. This

is because the PAPI questionnaire contained over 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 a question. There were 30 occasions when the PAPI interviewer had to refer back to a previous answer to determine the next question to be asked. There were nine questions that required the interviewer to perform an arithmetic calculation. All of these time-consuming and error-prone operations were done automatically, instantly, and correctly by the CAPI software.

### Missing Data

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

Research supports this. For example, a small-scale CAPI feasibility study conducted by Research Triangle Institute in 1987 found that over 90% of the errors made by PAPI interviewers were failures to record an answer; CAPI interviewers made no such errors (Sebestik et al., 1988). Groves and Mathiowetz (1984) demonstrated that CATI interviewers were able to follow complex skip logic virtually flawlessly, while PAPI interviewers made about five times as many errors. Presumably the favorable results obtained using CATI can also be obtained using CAPI.

The NLS/Y analysis looked at the three kinds of missing data— incorrect skips, refusals, and don't know responses. We compared the average rate of incorrect skips, refusals, and don't knows for the first ten rounds of the NLS/Y, carried out using PAPI, with similarly computed means for rounds 11 and 12 but using only the data collected with CAPI. While CAPI has no apparent effect on refusals and don't know responses, it completely eliminates illegal skips which account for the overwhelming proportion of missing data in the NLS/Y. Based on PAPI, about 1% of the total data points of responding households in the first 10 rounds were missing due to illegal skips. This rate was reduced to 0 in the CAPI experimental panels of Rounds 11 and 12. The reason is that the CAPI software ensured that the interviewer always followed the correct skip pattern.

### 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 than 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 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 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 is that respondents perceive computer-assisted interviews as being more confidential and therefore are more likely to report truthfully. Unfortunately, the Round 11 questions on drinking were not repeated in Round 12.

Some of the Round 12 experimental data did support the hypothesis that CAPI respondents are more willing to disclose sensitive information. 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 1 compares PAPI and CAPI responses to the two birth control questions separately for males and females. Frequencies on these items suggest some mode effect. Sixty-six percent of the male CAPI respondents report having used some form of birth control in the previous month compared to 58.5 percent for controls. When males reporting having used a method in the last month were asked how often they used it, 89.5 percent of the CAPI respondents reported that they always used it compared to 93 percent for the controls (Tables 3 and 4). While the differences in both cases are

small, they are statistically significant. On the question about specific methods, there were no clear differences between modes. Seventy-two percent of the female CAPI respondents report having used birth control in the last month as compared to 68 percent of the controls. This difference is statistically significant. Unlike males, females show no reporting differences across mode when asked how often. Similar to the males, there are no apparent differences in reports on the specific methods used.

In addition to these questions on birth control, females also are asked a large number of questions on current and former pregnancies, including questions about health-related behavior such as smoking, drinking, and drug use during pregnancy. We did not detect any mode differences in these items, although for many of these items 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 partially born out in 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 interview 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 five percent favoring paper and pencil. The remaining 48 percent felt that the two modes were equally confidential.

#### **Other Mode Differences in Response**

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 items included 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. Out of the 385 categorical variables, 48 (12.5%) gave rise to statistically significant CAPI-PAPI differences at the .05 level between CAPI and PAPI based on chi-square tests of independence. Out of the 60 continuous variables, 12 (20%) gave rise to signifi-

cant CAPI-PAPI differences at the .05 level of significance based on two-sample t-tests adjusted for unequal variances between groups. The overall percent of significant CAPI-PAPI differences is 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 CAPI-PAPI differences, we trimmed outlying values of the response 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. 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.

We found that most significant mode differences occurred in questions where the PAPI field interviewer 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 kind of error.

Table 2 presents four examples of significant CAPI vs. PAPI differences, examples which illustrate patterns in the significant mode differences we found. For each example, Table 2 shows the set of eligible respondents, the response category (i.e., the category that gave rise to the largest modal difference), the PAPI and CAPI percents responding in the category, and, in parentheses, the "Base n", i.e., the number of eligible CAPI or PAPI respondents who answered the question. Due to skip patterns and item nonresponse, Base n's are smaller than the corresponding numbers of complete cases. The number of complete cases equals 2306 for PAPI and 2287 for CAPI.

1. Household roster: CAPI had a small but significantly higher percentage of spouses listed in the household roster, about 54% for CAPI vs. 51% for PAPI. The greater percent of spouses recorded

in CAPI is probably due to the comprehensive electronic updating of the roster for additions, departures, and other corrections that is afforded by the CAPI software. This replaces an error-prone manual procedure for updating the roster in PAPI.

**2. Respondent's activity:** Perhaps the most important mode difference occurred in responses to a question on the respondent's activity last week. This NLS/Y question corresponded exactly to Q19 of the pre-1994 CPS questionnaire. CAPI shows a slightly smaller percent reporting "With a job but not at work," while PAPI shows a slightly smaller percent reporting "Other." We think these differences are due to central office editing of "Other" responses in PAPI, editing which did not occur in CAPI.

**3. Earnings:** Relatively more PAPI respondents chose to report their earnings on an hourly as opposed to a weekly or annual basis. In PAPI, but not in CAPI, a subsequent question on whether the respondent is paid by the hour appears on the same page of the questionnaire. It seems likely that some PAPI interviewers anticipated the subsequent question and obtained and entered all the hourly information at once, allowing them to skip over the subsequent question.

**4. Income:** On average, CAPI respondents reported slightly higher incomes in 1989 than PAPI respondents. The difference in mean incomes is on the borderline of statistical significance at the .05 level. There is no statistically significant difference when respondents who reported no income are dropped from the analysis. So the difference appears to be due to the slightly higher percent of PAPI respondents who reported receiving no income. Some respondents might report no income to avoid having to give a correct answer, and, if so, this result provides further evidence that CAPI respondents are more likely to disclose sensitive information.

### Survey Costs

Reducing survey costs is often cited as a likely benefit of CAPI. In theory at least, CAPI eliminates the need for all of the post-interview processing required by paper-and-pencil interviewing, thereby saving the not insignificant costs of keying, machine-editing, and programming to set up and perform these tasks. Skeptics argue that any savings in post-processing are offset or even exceeded by the increased cost of interviewing and by pre-field costs such as software design, increased training, and hardware acquisition.

Unfortunately, research on CAPI has not always

focused on cost, preferring instead to look at overall feasibility, respondent acceptance, interviewer acceptance, and so forth. One exception is the RTI test reported by Sebestik, et al. (1988) in which the costs of fielding the survey with CAPI were higher than with PAPI (even though hardware and software costs were excluded), largely because the hoped-for savings in processing did not materialize while the cost of interviewing rose. The experience on NLS/Y was somewhat different. The NLS/Y training costs ran about 70 percent higher for CAPI than for PAPI; this compares to about an 18 percent increase reported by RTI. The actual field interviewing costs on NLS/Y were about 12 percent higher, compared to 17 percent for RTI. Unlike the RTI experience, however, the NLS/Y data processing costs—including the hardware—were much lower for CAPI than for PAPI, although the overall cost per case of the NLS/Y CAPI cases was about 12 percent higher than cost per case of PAPI cases.

In both the RTI test and the NLS/Y experiment, the higher CAPI field interviewing costs were not due to increased interviewing time, but rather to an added support burden the transition to CAPI required and, in the case of NLS/Y, allocation of hardware costs to the field effort rather than to data processing. As noted above, the average NLS/Y interview length was much shorter with CAPI than with PAPI.

On balance, the higher pre-interview costs of CAPI in NLS/Y, including the costs of portable computers and software development, were largely offset by the virtual elimination of post-interview data processing. The overall per unit cost of CAPI cases was only slightly higher than the overall per unit cost of PAPI cases.

### Conclusions

Compared with PAPI, CAPI yields lower item nonresponse rates, because of the elimination of missing data due to incorrect skips. 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, can be substantially less with CAPI, especially if the questionnaire requires branching based on retrieval of previous responses. Some respondents view computer-assisted interviews as more confidential than traditional face-to-face methods and may be more forthcoming in response to sensitive questions.

We also found that higher pre-interview costs of

CAPI, including the costs of portable computers and software development, were largely offset by the virtual elimination of post-interview data processing. 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.

We need to be cautious in generalizing the NLS/Y experience to other surveys. The NLS/Y incurs very large data processing costs (relative to other costs) by virtue of a good deal of materials generation— assignment lists, complex facesheets, locating information, and so forth— prior to field and an extremely rigorous post-data-entry batch cleaning. As we have just seen, the questionnaire is very complicated, the sort that can be administered more efficiently by computer than by traditional PAPI methods. More generally, we also should 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.

## References

Baker, R., 1990. What We Know About CAPI: Advantages and Disadvantages. 1990 American Association of Public Opinion Research Conference. Lancaster, PA, 17-20 May.

Baker, R. and Bradburn, M., 1991. CAPI: Impacts on Data Quality and Survey Costs, Paper presented at the 1991 Public Health Conference on Records and Statistics, Washington, D.C.

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

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

Bradburn, N., Frankel, M., Hunt, E., Ingels, J., Schoua-Glusberg, A., Wojcik, M. and Pergamit, M., 1991. A Comparison of Computer-Assisted Personal Interviews with Personal Interviews in the National Longitudinal Survey of Labor Market Behavior—Youth Cohort. Proceedings of the Annual Research Conference, Washington, D.C.: U.S. Bureau of the Census.

Bradburn, N., Frankel, M., Baker, R., and Pergamit, M., 1991. 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.

Dippo, C., 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. and Mathiowetz, N., 1984. Computer-Assisted Telephone Interviewing: Effects on Interviewers and Respondents. Public Opinion Quarterly 48:356-369.

Keller, W., Metz, K. and Bethlehem, J., 1990. The Impact Of Microcomputers on Survey Processing at the Netherlands Central Bureau of Statistics. In Proceedings of the U.S. Bureau of the Census Annual Research Conference. Washington, DC: U.S. Bureau of the Census.

Lyberg, L. and Dean, P., 1989. Plans for Computer Assisted Data Collection at Statistics Sweden." Proceedings of the 45th Session of the International Statistics Institute, Book III, Topic 18.2.

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

Nicholls, W. and Groves, R., 1986. The Status of Computer-Assisted Telephone Interviewing: Part I. 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.

Sebestik, J.H., Zelon, H., DeWitt, D., O'Reilly, J. and McGowan, K., 1988. Initial Experiences with CAPI. In Proceedings of the Fourth Annual Research Conference. Washington, DC: U.S. Bureau of the Census.

Speizer, H. and Dougherty, D., 1991. Automating Data Transmission and Case Management Functions for a Nationwide CAPI Study. In Proceedings of the U.S. Bureau of the Census Annual Research Conference, Washington, D.C.: U.S. Bureau of the Census.

Waterton, J. and Duffy, J., 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.

Table 1. Responses to Questions on Birth Control

QUESTION	RESPONSE	MALES		FEMALES	
		PAPI %	CAPI %	PAPI %	CAPI %
During the last month have you or your partner/spouse used any form of birth control? (n=4067)	Yes	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	93.0%	89.5%	95.0%	94.7%

Table 2. Examples of Variables with Significant CAPI-PAPI Differences

VARIABLE	ELIGIBLE	CATEGORY	PAPI % (Base n)	CAPI % (Base n)
<b>1. HOUSEHOLD ROSTER QUESTION:</b>				
R's spouse on HH roster? (FI coded)	All	Yes	50.7% (2270)	54.1% (2280)
<b>2. LABOR FORCE QUESTION:</b>				
R's activity last week (CPS Q19)	All	With a job but not at work	5.1% (2254)	3.0% (2271)
	All	Other	4.3% (2254)	6.5% (2271)
<b>3. EARNINGS QUESTIONS:</b>				
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)
<b>4. INCOME QUESTION:</b>				
R's income in 1989	All	Range: \$0 to \$99999	\$15154 (2240)	\$15950 (2247)