

EVALUATION OF THE USE OF CATI IN THE CURRENT POPULATION SURVEY

Harland H. Shoemaker, John M. Bushery, Lawrence S. Cahoon
U. S. Bureau of the Census, Washington, D.C. 20233

1. INTRODUCTION

This paper continues a discussion of Census Bureau research using CATI in the Current Population Survey (Bushery 1987). Based on this research, the Census Bureau has begun a limited and carefully controlled introduction of the CATI methodology into the survey (Walsh 1988b).

The Census Bureau conducts the Current Population Survey (CPS) for the Bureau of Labor Statistics. The survey provides estimates of various labor force characteristics, particularly employment and unemployment.

The advantages of CATI have been covered in numerous papers (Fink, 1983; Groves, 1983; Coulter, 1985; Tortora, 1985; Nicholls, 1986; Groves, 1986; Bushery, 1987). The Census Bureau is interested in CATI for all the usual reasons: more efficient survey management, faster data transmission and processing, better supervision and training. CATI also promises the possibility of reducing the Census Bureau's staffing problems in "hard-to-recruit" areas. Fewer field representatives (interviewers) are needed if CATI interviewers handle a sizable portion of the CPS workload.

CATI will not increase telephone interviewing in the CPS. Over half of all CPS cases are interviewed by telephone. CATI will shift the work from the field representatives (FRs) who conduct telephone interviews in their homes to the interviewers in the centralized CATI facility.

Before introducing a change of methodology into its surveys, the Census Bureau carries out careful evaluations to measure any effects associated with the change. Of major importance is the need to measure changes in the survey's time series caused by the new methodology.

The CPS CATI evaluation began in June 1985. The second phase of the evaluation continued from August 1986 through December 1988. Bushery et al. (1987) describes the design and methodology of this study.

The CATI Test sample consisted of 3,000 housing units per month, divided equally among 30 metropolitan statistical areas (MSAs). These MSAs were selected randomly from a frame of 109 eligible areas. The regular CPS sample provided the Control group for the CATI evaluation. The monthly Control sample consisted of about 9,400 units. Control sample sizes in the individual MSAs ranged from 100 to 1,200 units, depending on the CPS sample size in the MSA. The Test sample used an abbreviated rotation scheme: four months in sample, then retirement from the study. The Control sample used only the corresponding interviews from the CPS's 4-8-4 rotation scheme, the first four interviews. This let us phase in the study sample in only four months.

2. RESPONSE RATES

2.1 Unit Response

High response rates and low item nonresponse are important in controlling nonsampling error. We classify sample cases as completed interviews or as type A, B, or C

noninterviews. Type As are eligible units which are not interviewed. Respondent refusal, no one home, and temporary absence are the main reasons for type As. Type B and C noninterviews are cases not eligible for the survey. Type Bs are vacant or otherwise out of scope and type Cs are not housing units.

Because CPS response rates are very high, we worried that using CATI might hurt them. Respondents may be less willing to cooperate with unfamiliar interviewers calling from the CATI site. The CATI interviewers are not evaluated on response rates and might expend less effort trying to complete a difficult interview. Finally, the short interview period allows local FRs little time to follow up nonresponse cases "recycled" from the CATI facility to the field.

The differences in response rates between the Test and Control samples are small and cause no concern. Table 1 reveals significant differences only for type B cases, primarily vacant units.

During the first month-in-sample (MIS 1) there were no operational differences between the Test and Control groups. Unexpectedly, the Test group's type A and refusal rates were lower than the Control's: 6.4 percent versus 7.4 percent, and 2.6 percent versus 3.0 percent, respectively. During MIS 2-4 when the procedures differed, the response rates showed no significant differences. These results lead us to conclude the mixed mode CATI methodology probably has no adverse impact on response rates.

2.2 Item Response

We evaluated only items with relatively high nonresponse and items critical to determining labor force status.

The most notable difference between a paper questionnaire and CATI is the automated control of question flow in the CATI environment. Some differences in how the two instruments are implemented in the CPS may cause different response rates. For example, the paper questionnaire provides for "refused" responses only for the two questions on hourly wages and weekly earnings. This option is not available for any other questions. The CATI instrument, in contrast, allows both refusal and "don't know" entries for all items.

We thought that differences between field and CATI supervisory practices may affect item response rates. Item response rates are included in FR performance ratings, but are not used to evaluate CATI interviewers. Further, the CATI system does not permit callbacks to collect data for missing items. These factors would tend to lower CATI item response rates. However, CATI supervisors can monitor interviewer performance more easily and provide prompt feedback when there are problems. We can easily revise the CATI instrument to probe for items with low response. In fact, early in the study, CATI nonresponse was too high for some items (Table 2a). We reduced item nonresponse in CATI to an acceptable level by adding probes and providing feedback on item nonresponse to the CATI managers. There are still some

items where the CATI response rate is lower than the Control's, but the difference is negligible. For the difficult hourly wage question, the response rate is now actually better for the CATI methodology (Table 2b).

3. POPULATION COVERAGE

In the CPS some population groups have significant undercoverage. Most notably, we miss 25 percent to 30 percent of the young black males in the population (Shapiro 1988). Because methodologies with different coverage can produce different estimates, we paid particular attention to coverage in this study.

We examined three areas of coverage: coverage of housing units, overall coverage of persons, and coverage of persons within housing units. Preliminary analyses revealed two issues.

The Test group estimate of eligible housing units was lower than the Control group estimate. Also, the Test group estimate of black males declined significantly over month-in-sample.

3.1 Coverage of Population and Housing

The Test group's estimates of eligible housing units (HUs) and population were consistently lower than the Control's. Close examination suggested that a higher vacancy rate in the Test group caused the difference. Table 3 displays housing and population estimates for the study. Neither the population estimates nor within-unit coverage showed significant differences between procedures.

We believe sampling variability probably caused much of the apparent difference in housing estimates. It appears that the Test sample hit neighborhoods with higher vacancy rates than the Control. A comparison of vacancy rates and housing estimates between the Test group and a subset of the Control falling in the same neighborhoods tends to support this conclusion. Neither the Test group vacancy rate nor housing estimate displays significant differences from the Control's in the same neighborhoods (Walsh 1988a).

Part of the coverage difference is due to a small loss of sample cases in the CATI recycling operation. The losses occur when cases are recycled from the CATI facility to the field. A few cases simply "drop out" of the system. Over the life of the study, the system lost about 5 percent of the cases recycled from the CATI facility to the field. This amounted to approximately 0.5 percent of all Test cases.

The system relied on clerical intervention to move cases from the CATI site to the field and back again. Troubleshooting revealed that clerical updating in the control system was a major factor in losing recycled cases. We are in the process of installing an automated data transfer system. This will eliminate the clerical updating.

3.2 Black Estimates

The Test group's estimates of black males for MIS 2-4 averaged about 4.5 percent lower than the estimates for MIS 1 (Table 4). The estimates for each MIS decrease as the sample ages. Neither the Control group nor the regular CPS exhibits this phenomenon (Walsh 1988b).

Special analyses suggest the Test procedure may have lost persons from interviewed units. There is no evidence that the Test group picks up fewer movers-in than the Control. Three of the 30 sample areas account for 72 percent of the net loss of sample persons. The small numbers involved in this problem make it very difficult for us to pin down the causes of the decrease, but in these three areas the reason for the loss in black males was almost equally distributed among lost HUs containing black males, lost persons within interviewed units and noninterviews of previously interviewed HUs.

4. LABOR FORCE CHARACTERISTICS

4.1 Civilian Labor Force and Unemployment Rates

The overall Test group civilian labor force (CLF) rate was significantly higher than the Control's. Significant differences in the CLF rate were apparent only for total males and white males (See Table 5). Figure 1 plots the monthly CLF rate differences between the Test and Control.

In November 1987, the Test group CLF rate jumped about 2 percentage points. The difference between Test and Control increased accordingly. By February 1988, the rate and the differences returned to normal levels. Our investigations suggest sampling variability accounts for this temporary jump in the Test group CLF rate. Overall, Figure 1 shows the Test group's estimated CLF rate to be consistently higher than the Control's. This pattern ordinarily would be highly significant, statistically. However, the 75 percent overlap between monthly samples weakens this conclusion.

Figure 2 suggests that using CATI may increase CPS unemployment rates. The Test group's estimated unemployment rate averaged a half percentage point higher than the Control's, but this is not a statistically significant difference. The monthly unemployment rates, like the CLF rates, are consistently higher in the Test group than the Control. Twenty-two of the 25 monthly Test group estimates are higher than the Control's. If it weren't for the 75 percent overlap between monthly samples, this would be a highly statistically significant result.

4.2 NILF Estimates

Estimates of persons not in the labor force (NILF) are significantly lower for the Test group, for both males and females, total persons and whites. These lower Test group NILF estimates are related to the higher CLF rates. We analyzed microdata files to gain some insight into this phenomenon, but found no explanation.

4.3 Rotation Group Bias Patterns

It is widely known that CPS estimates are affected by rotation group bias (Bailar 1975). This means that estimates from different rotation groups have different expected values. We examined the effect of CATI on rotation group bias by comparing average CLF and unemployment rates by MIS for the two methodologies.

For the CLF rate, we found that the Test and Control group exhibit different patterns of rotation group bias. The Test group estimates are more stable across MIS, suggesting that differences in this bias are

somewhat smaller using CATI. The unemployment rate shows no significant difference in rotation group bias patterns between Test and Control. We conclude that the CATI methodology may improve our estimate of the CLF rate by stabilizing rotation group bias.

5. RESPONSE VARIANCE

Response error is an important factor in data quality. Throughout 1988, we conducted CATI reinterviews to measure simple response variance in the labor force estimates of the CPS/CATI study.

The reinterview sample consisted of one eighteenth of the households interviewed at the CATI facility. We conducted reinterviews at about 87 percent of the 793 cases so selected.

We examined two measures of response variance, the gross difference rate (GDR) and the index of inconsistency. The GDR is the proportion of all cases which reported a particular category in the original interview or the reinterview, but not in both. The simple response variance for a category is one half the GDR. The index of inconsistency is the ratio of a category's simple response variance to its total variance.

The following rule of thumb helps to interpret the index of inconsistency. Indices below 20 are low, indicating response variance is not a major problem. Indices between 20 and 50 are moderate, suggesting there are some problems with inconsistent reporting. Values of the index above 50 suggest the responses are not reliable. When the index is high, the methods used to collect the data probably need improvement or the concept itself may not be measurable.

The CPS/CATI reinterview was not designed to be comparable to the CPS reinterview. Nevertheless, the measures of simple response variance in this study approximate those in the regular CPS. The index of inconsistency is low for employed (6.7) and NILF (8.9) and moderate for unemployed (38.1) as estimated by CATI.

Although the gross difference rate for unemployed is an apparently low 2.4 percent, the index of inconsistency shows that response variance accounts for about 38 percent of the total variance on the estimate.

6. SUMMARY

The CATI methodology produces results comparable to those from the regular CPS for most of the characteristics we examined: response rates, item nonresponse, coverage, labor force characteristics and response variance.

There are notable exceptions, namely coverage of black males and estimates of the CLF rate. There is some indication that the unemployment rate also is affected.

The differences in labor force estimates do not make CATI undesirable. Different is sometimes better. We believe that is the case here. However, the loss of black males in the Test group does disturb us. We are planning to continue our analysis of the data to further isolate the causes.

REFERENCES

Bailar, Barbara A. (1975), "The Effects of Rotation Group Bias on Estimates from

Panel Surveys," Journal of the American Statistical Association, Vol. 70, No.349, p. 23-30.

Bushery, J., L. Weidman, S. Briley, and S. Ciocchetto (1987), "Using CATI in the Current Population Survey," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Coulter, Richard (1985), "A Comparison of CATI and NonCATI on a Nebraska Hog Survey." Staff Report No. 85, Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture.

Ferrari, Pamela W. (1984), "Preliminary Results from the Evaluation of the CATI Test for the 1982 National Survey of Natural Scientists and Engineers." Unpublished research report, U.S. Bureau of the Census.

Ferrari, Pamela W. (1986), "An Evaluation of Computer Assisted Telephone Interviewing Used During the 1982 Census of Agriculture." Unpublished evaluation report, Agriculture Division, U.S. Bureau of the Census.

Fink, James C. (1983), "CATI's First Decade: The Chilton Experience." Sociological Methods and Research, Vol. 12, No. 2, p. 153-168.

Groves, Robert M. (1983), "Implications of CATI: Costs, Errors, and Organization of Telephone Survey Research." Sociological Methods and Research, Vol. 12, No. 2, p. 199-215.

Groves, Robert M., and William L. Nicholls II (1986), "The Status of Computer Assisted Telephone Interviewing, Part II - Data Quality Issues." Journal of Official Statistics, Vol. 2, No. 2, p. 117-134.

Nicholls, William L., II and Robert M. Groves (1986), "The Status of Computer Assisted Telephone Interviewing, Part II - Data Quality Issues." Journal of Official Statistics, Vol. 2, No. 2, p. 93-116.

Rustemeyer, Anitra, Gerald H. Shure, Miles S. Rogers, and Robert J. Meeker (1978), "Computer Assisted Telephone Interviewing: Design Considerations." Proceedings of the Section on Survey Research Methods, American Statistical Association, p. 1-8.

Shapiro, Gary M. and Donna L. Kostanich (1988), "High Response Error and Poor Coverage are Severely Hurting the value of Household Survey Data." Proceedings of the Section on Survey Research Methods, American Statistical Association.

Tortora, Robert D. (1985), "CATI in an Agricultural Statistical Agency." Journal of Official Statistics, Vol. 1, No. 2, p. 301-314.

Walsh, Thomas, et.al. (1988a), "Preliminary Evaluation of CATI Data Quality and Cost in the Current Population Survey," April 11, 1988. Unpublished research report, U. S. Bureau of the Census.

Walsh, Thomas, et. al. (1988b), "Evaluation of CATI Data Quality and Cost in the Current Population Survey, CATI Research Report No. CPS-4," October 27, 1988.

ACKNOWLEDGEMENTS

The authors wish to thank the following persons for their contributions to this report: William Nicholls, David Chapman, Kathleen Creighton, Irwin Schreiner, Lorie Baker, Carol Gunlicks, and Timothy Downey. We would also like to thank Kathy Cochran and Sue Chandler for their skill and patience in typing this paper.

Table 1: Test and Control Average Response Rate Statistics
CPS/CATI Phase II, January 1987-October 1988

Total Month-in-Sample	Test	Control	Difference	Error
Total Units	2939	8951	---	---
Eligible Units	2555 (86.93%)	7895 (88.20%)	-1.27%	.79
Ineligible Units	384 (13.07%)	1056 (11.80%)	1.27%	.79
Type B Noninterview-Total	351 (11.94%)	951 (10.62%)	1.32%*	.76
Type B--Vacants	244 (8.30%)	645 (7.21%)	1.09%*	.65
Type C Noninterviews	33 (1.12%)	104 (1.16%)	-0.04%	NA
Eligible Units	2555	7895	---	---
MIS 1	644	1975	---	---
MIS 2-4	1911	5922	---	---
Interviewed Households	2414 (94.48%)	7448 (94.34%)	0.14%	.19
Type A Noninterviews	141 (5.52%)	447 (5.66%)	-0.14%	.19
MIS 1	41 (6.37%)	146 (7.39%)	-1.02%**	.27
MIS 2-4	99 (5.18%)	303 (5.12%)	0.06%	.22
Type A--Refused	78 (3.05%)	244 (3.09%)	-0.04%	.15
MIS 1	17 (2.64%)	59 (2.99%)	-0.35%**	.17
MIS 2-4	60 (3.14%)	186 (3.14%)	0.00%	.17

Table 2a: Test (Total), Test (CATI), and Control Average Item Response Rates
CPS/CATI Phase II, January 1987-June 1987

Item	Item Response Rates (%)				S.E.
	Test (Total)	Test (CATI)	Control	Difference ¹	
Hours worked	98.49	97.67	99.75	-1.26 ***	0.21
Reasons for working under 36 hours	97.24	96.45	98.66	-1.42 **	0.56
Job seeking methods Available for work	98.35	97.52	99.61	-1.25	0.79
Hourly wage	98.70	98.33	99.06	-0.36	0.70
Weekly earnings	80.01	80.01	86.09	-6.08 ***	1.95
	81.14	81.49	81.40	-0.26	2.16

Table 2b: January 1987-June 1988

Item	Item Response Rates (%)				S.E.
	Test (Total)	Test (CATI)	Control	Difference ¹	
Hours worked	99.20	98.83	99.80	-0.60 ***	0.09
Reasons for working under 36 hours	98.13	97.64	98.72	-0.59 **	0.27
Job seeking methods Available for work	98.80	97.92	99.56	-0.76 *	0.41
Hourly wage	98.20	97.87	99.01	-0.81	0.50
Weekly earnings	86.65	88.53	84.65	2.00 **	0.98
	82.29	84.36	81.50	0.79	0.84

¹ Difference = Test (Total)-Control

* Significant at alpha = 0.10

** Significant at alpha = 0.05

*** Significant at alpha = 0.01

Table 3: Test and Control Estimates of Eligible Housing Units
Persons Age 16 and Over, and Persons Per Household
CPS/CATI Phase II, November 1986 - November 1988

Monthly Average Estimates	Eligible HUs (1000s)	Persons 16+ (1000s)	Persons Per HU
Test	25082	51486	2.05
Control	25936	53251	2.05
Difference	-854*	-1766	.00
SE(Diff.)	514	1304	.02

* Significant at alpha = 0.10

Table 4: Test Estimates of Blacks (16+) by MIS
CPS/CATI Phase II, November 1986 - November 1988

Month-in Sample (MIS)	Test Group			
	Black Males		Black Females	
	Estimate (1000s)	Difference from MIS 1	Estimate (1000s)	Difference from MIS 1
MIS 1	804	--	1089	--
MIS 2	791	-13	1095	6
MIS 3	776	-28	1066	-23
MIS 4	736	-68	1039	-50
AVG. MIS 2-4	768	-36	1067	-22
% Diff.		-4.48 ***		-2.02
SE (Diff.)		14.51		17.61

*** Significant at alpha = 0.001

Table 5: CPS CATI Phase II Labor Force Estimates
November 1986 To November 1988

Monthly Average Estimates	Total Persons 16+ (1000s)	Civilian Labor Force (1000s)	Em- ployed (1000s)	Unem- ployed (1000s)	Not in Labor Force (1000s)	CLF Rate	UE Rate
Total Persons							
Test Group	51486	34802	32610	2192	16684	67.59	6.30
Control Group	53252	35388	33317	2071	17864	66.45	5.85
Difference	-1766	-586	-707	121	-1180***	1.14**	0.45
S.E. (Diff.)	1304	1003	978	130	428	0.54	0.37
Male							
Test Group	24019	18750	17614	1137	5269	78.06	6.06
Control Group	25034	19233	18106	1127	5802	76.83	5.86
Difference	-1015	-483	-492	10	-533***	1.23**	0.20
S.E. (Diff.)	668	567	551	69	191	0.63	0.36
Female							
Test Group	27466	16051	14996	1055	11415	58.44	6.57
Control Group	28217	16155	15211	944	12062	57.25	5.85
Difference	-751	-103	-215	111	-647**	1.19	0.72
S.E. (Diff.)	683	500	486	75	303	0.76	0.46
White							
Test Group	41890	28485	27090	1395	13405	68.00	4.90
Control Group	44047	29369	27968	1401	14677	66.68	4.77
Difference	-2157	-884	-879	-6	-1273***	1.32*	0.13
S.E. (Diff.)	1454	1105	1070	72	474	0.69	0.22
White Male							
Test Group	19841	15668	14901	767	4173	78.97	4.89
Control Group	20988	16295	15498	797	4693	77.64	4.89
Difference	-1147	-627	-596	-30	-520***	1.33**	0.00
S.E. (Diff.)	743	618	602	42	179	0.59	0.25
White Female							
Test Group	22048	12817	12188	628	9232	58.13	4.90
Control Group	23059	13074	12471	604	9985	56.70	4.62
Difference	-1010	-257	-282	25	-753**	1.43	0.28
S.E. (Diff.)	749	530	511	46	327	0.94	0.32
Black							
Test Group	7276	4739	4067	672	2537	65.13	14.18
Control Group	7054	4559	3975	585	2495	64.63	12.82
Difference	221	180	92	87	42	0.50	1.36
S.E. (Diff.)	743	468	380	114	309	1.60	1.54
Black Male							
Test Group	3059	2237	1926	311	822	73.12	13.89
Control Group	3028	2157	1877	280	871	71.24	13.00
Difference	32	80	50	30	-48	1.88	0.89
S.E. (Diff.)	323	226	186	55	121	2.07	1.66
Black Female							
Test Group	4216	2502	2140	361	1714	59.34	14.44
Control Group	4027	2402	2098	304	1625	59.66	12.67
Difference	190	100	43	57	90	-0.32	1.77
S.E. (Diff.)	432	252	204	66	204	1.79	1.76

* significant at alpha = 0.10

** significant at alpha = 0.05

*** significant at alpha = 0.01

Figure 1: Differences in CLF Participation Rates

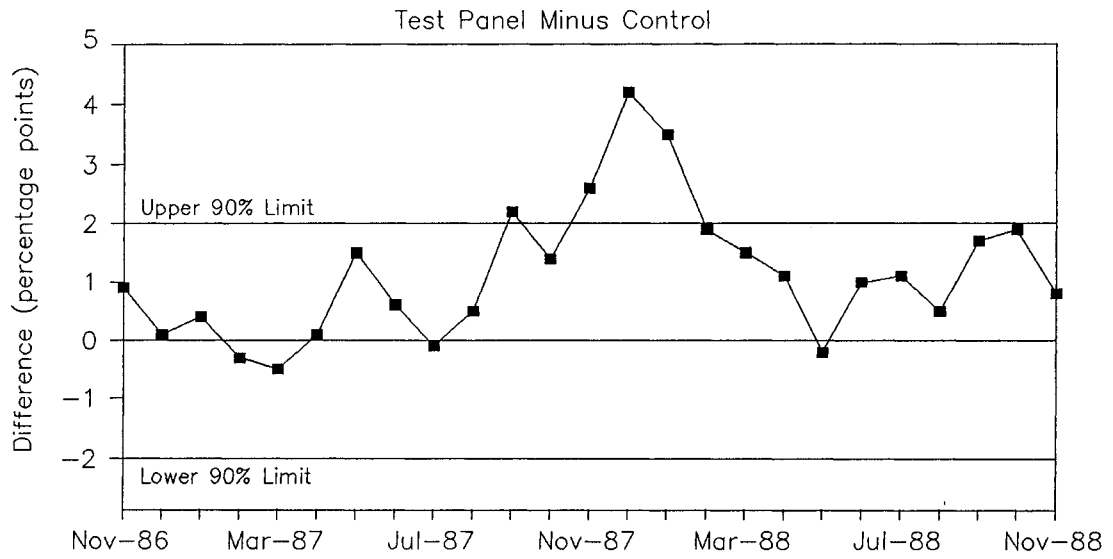


Figure 2: Differences in Unemployment Rates

