

RESEARCH REPORT SERIES
(*Survey Methodology* #1986-03)

CATI Phase II CPS Data Analysis Plan

John Bushery
Easley Hoy
Jennifer Rothgeb
Lynn Weidman
Glenn White

Center for Behavioral Science Methods
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report issued: March 2021

Disclaimer: This report is released to inform interested parties of research and to encourage discussion of work in progress. Any views expressed are those of the authors and not those of the U.S. Census Bureau.

Abstract

The primary objective of the CATI CPS Data Analysis Project is to evaluate CATI against the current CPS methodology with regard to labor force estimates, response rates, and gross change. The gross change comparisons will provide an indirect measure of response variability in CATI relative to the CPS. Other objectives include comparisons of item nonresponse, coverage, and quality of industry and occupation data. We also hope to track respondent characteristics and analyze households by telephone availability status. The estimates used in these analyses will be obtained from the sample areas represented in the CATI design. A detailed discussion of how we plan to meet these objectives is provided below.

Keywords: CPS, mode, labor force

Suggested Citation: John Bushery, Easley Hoy, Jennifer Rothgeb, Lynn Weidman, Glenn White (1986). **CATI Phase II CPS Data Analysis Plan.** *Research and Methodology Directorate, Center for Behavioral Science Methods Research Report Series (Survey Methodology #1986-03)*. U.S. Census Bureau. Available online at <https://www2.census.gov/adrm/CBSM/sm1986-03.pdf>



CATI Phase II CPS Data Analysis Plan



John Bushery
Easley Hoy
Jennifer Rothgeb
Lynn Weidman
Glenn White

CATI PHASE II
CPS DATA ANALYSIS

-CONTENTS-

- I. GOALS AND OBJECTIVES

- II. DETAILED OBJECTIVES
 - A. Major Labor Force Estimates
 - B. Response Rates and Item Nonresponse
 - C. Response Variance
 - D. Coverage
 - E. Characteristics of Respondents
 - F. Quality of Industry and Occupation Data
 - G. Telephone Availability

- III. SCHEDULE AND PRIORITIES

Attachments: Table Layouts

Analysis Plan
CATI CPS Data Analysis Team

I. INTRODUCTION

The primary objective of the CATI CPS Data Analysis Project is to evaluate CATI against the current CPS methodology with regard to labor force estimates, response rates, and gross change. The gross change comparisons will provide an indirect measure of response variability in CATI relative to the CPS. Other objectives include comparisons of item nonresponse, coverage, and quality of industry and occupation data. We also hope to track respondent characteristics and analyze households by telephone availability status. The estimates used in these analyses will be obtained from the sample areas represented in the CATI design. A detailed discussion of how we plan to meet these objectives is provided below.

The 30 metropolitan areas included in the CATI Phase II Study were selected from a frame of 109 metropolitan areas. To be eligible for the CATI Study an area had to be metropolitan, consist of one or more self-representing CPS PSUs, and provide a sample of 70 or more housing units (HUs) to the CPS. Because the initial implementation of CATI in the "live" CPS probably will be in multi-interviewer metropolitan PSUs, we required a minimum sample size of 70 HUs. This eliminated single-interviewer PSUs from the frame of eligible areas.

Comparisons between CATI and the "regular CPS" will be made only for the 30 sample areas. Managers of the CPS in SMD and DSD have decided that estimates for the entire frame of CATI eligible areas are of marginal utility and do not justify the resources required for their production.

The CATI sample design differs from that of the regular CPS in a number of respects. First, the CATI sample will be divided into only four rotation groups. Each CATI rotation group will be interviewed for 4 consecutive months, then retired from the study. This rotation scheme allows for a rapid phase-in of the CATI sample. In an effort to maintain comparability, only the first four months-in-sample (MIS) of the CPS will be used in this study. Also, problems with staff availability precluded selection of new construction sample (i.e., permit segments) for CATI, so only old construction sample (address, area, and special place segments) from the regular CPS will be used in this study. The CPS sample just described (i.e., old construction only, MIS 1-4 only) for the 30 sample areas only will comprise the control sample for the study. We will refer to this sample as the "CPS control sample" or the "Control sample" throughout this document.

It must be mentioned that the CATI methodology under consideration in this study is not a "pure CATI" methodology. Rather it approximates the type of CATI system which would be implemented in the CPS on a production basis. This system involves a mixture of personal visit, CATI, and decentralized non-CATI telephone interviews. One fourth of the CATI sample always is interviewed in person (the incoming rotation group). Of the remaining three rotation groups, only about three-fourths will be interviewed by the CATI facility. The remainder will be retained by the Field offices as unsuitable for CATI (for example, no telephone) or will be "recycled" to the field because contact could not be made via CATI.

Finally the estimation procedure for labor force characteristics in this study is a simplified version of the CPS estimation procedure. Ratio estimation (the first and second stage ratio adjustments) and composite estimation will not be used in this study. Estimation for both the CATI and the Control samples will be performed using this simplified procedure, so the estimates for the two methodologies will be comparable.

II. DETAILED OBJECTIVES

We intend to compare labor force estimates, response rates, coverage, gross change in labor force status, quality of coding for industry and occupation data, and characteristics of respondents between the CATI methodology and the regular CPS methodology. We expect to find some differences between the two methodologies, so we also are interested in the magnitudes of the differences and where they occur.

A. Labor Force Comparisons

Our goal here is to determine whether the two methodologies appear to produce different estimates of labor force status. We plan to compare the distribution of weighted estimates of labor force status among the three categories: Employed, Unemployed, Not in Labor Force (NILF). We will test to determine whether the distribution produced by CATI differs from that produced by the Control. This analysis will not allow us to decide which methodology is superior, but only whether the two differ. We plan to use log linear analysis to compare these labor force distributions. We also will compare estimated levels and rates of unemployment and labor force participation. Finally, each month we will produce plots comparing key estimates for the two methodologies.

1. Hypotheses

We will test the hypotheses that the CATI and CPS labor force distributions are the same. We will make comparisons for the major demographic categories listed below.

Summary:	total persons 16+
by sex:	male, female
by race:	Black, White, other
by ethnicity:	Hispanic, not Hispanic
by age:	16-19, 20-24, 25+

Although data files will contain estimates for each of the 36 cross-classified demographic categories defined above, variability on these more detailed categories will be very high, so testing and tabulations will be provided for only the larger (i.e. marginal) demographic categories. Similarly, because of the small sample sizes and corresponding low reliability, estimates for individual CATI areas will not be tabulated.

2. Key Variables, Tabulations and Files Needed for Analysis

We will be able to obtain the desired labor force tabulations using available mainframe or PC software. A special data file will be produced, which contains tabulations of age/race/sex/ethnicity labor force status for each methodology, for six random groups in each sample area. This random group file will be used to estimate variances of the labor force data. In order to examine the data by month-in-sample, it will be necessary to replicate this file by MIS. This file can be used on the UNIVAC and downloaded for use on IBM-PCs. Attachment A. illustrates the tabulations we plan to produce.

3. Methodology, Formulas, Models

a. Analysis Model

Log linear analysis relies on a number of assumptions: sufficiently large samples are used, unweighted data are used, and the individual sample elements all are independent. Our analysis violates the second assumption because we plan to use weighted data. The third assumption also is violated because of correlation within CPS clusters

(segments). For simplicity we will begin our discussion as if all these assumptions are fulfilled. We plan to use software which makes adjustments to account for these violations. We are interested in the distribution of the vector of counts,

$$\underline{L} = (E, U, N)'$$

where E = employed
 U = unemployed
 N = not in labor force (NILF).

Our null hypothesis is,

Ho: The distribution of \underline{L} is the same for CATI and CPS.

Letting f_{ij} denote the observed frequency of labor force status j for interview methodology i , a log linear model represents $\ln(f_{ij})$ as the sum of specified effects. Under Ho the model is

$$\ln[E(f_{ij})] = \mu + \alpha_i + \beta_j$$

where the α_i represent effects due to interview methodology and the β_j represent effects due to labor force status. The cell frequencies, f_{ij} , are estimated levels for labor force status and are calculated as described in section 3.b. below.

This model allows for a labor force status difference between the two methodologies, but no interaction between method and labor force status. We also plan to use models which measure such interactions, as well as main effects for demographic status (race, sex, etc.).

There is computer software available (e.g. BMDP, SPSS, GLIM) which will allow us to fit log linear models using maximum likelihood techniques to estimate the effects parameters. Perhaps more useful for us is software developed by Bob Fay (CPLX), to perform log linear analysis. This program automatically makes the adjustments needed to account for using weighted data from complex sample designs. A table such as that shown below will be used in computing the test statistic for this analysis.

		<u>LABOR FORCE STATUS</u>		
INTERVIEW METHODOLOGY		Employed	Unemployed	NILF
		(1)	(2)	(3)
		β_1	β_2	β_3
CATI (1)	α_1	$f_{11} (\hat{f}_{11})$	$f_{12} (\hat{f}_{12})$	$f_{13} (\hat{f}_{13})$
CPS (2)	α_2	$f_{21} (\hat{f}_{21})$	$f_{22} (\hat{f}_{22})$	$f_{23} (\hat{f}_{23})$

The \hat{f}_{ij} 's are the maximum likelihood estimates of the f_{ij} for the given model. We can then compute the likelihood ratio goodness of fit test statistic,

$$G^2 = 2 \sum_{i,j} f_{ij} [\ln(\hat{f}_{ij})/\ln(f_{ij})] . \quad (1)$$

Under H_0 this statistic would have a Chi-Square distribution with 2 d.f. Because we will be using weighted data, we will make the appropriate adjust-

ments when comparing G^2 with its critical value. We plan to perform this test for monthly data as well as for cumulative data. In this way we can see how the distribution behaves from month-to-month.

To test for interaction effects between methodology and race, sex, age on labor force status we will fit models to three-way tables. For example, to determine if there is an interaction between methodology and race, we add effects for race and the race-by-methodology interaction and test for the interaction effect being zero. If it is, then race does not affect the distribution of L differently for CATI and the Control. The test statistic is the difference between the individual test statistics of the

form G^2 for the models with and without the interaction term.

One of the assumptions of this procedure is that sufficiently large samples are available. This is because the distributions of the test statistics are valid only asymptotically. We decided to use six random groups to assure sufficiently large cell sizes to permit the desired analyses.

b. Estimation of the Observed Cell Frequencies

The estimates of observed labor force frequencies (the f_{ij} in the log linear model) will be computed

as described below. A simplified weighting procedure will be used, with a single cell noninterview adjustment within each sample area and no first- or second-stage ratio adjustments or composite estimation. We will begin collecting data and making estimates in November 1986, when the CATI sample has been phased-in completely. For simplicity, this discussion covers only a single labor force category, for example, level unemployed. Let Y refer to the number of unemployed persons estimated by a methodology (CATI or the Control). In terms of the model, let

$$f_{ij} = \sum_{a=1}^{30} \bar{Y}_a \quad (2)$$

where

$$\bar{Y}_a = \sum_{k=1}^{n_a} W_{ak} \cdot \delta_{ak} \quad (3)$$

with

n_a = the number of sample persons in the a-th area

W_{ak} = the final weight for the k-th sample person in the a-th area.

This weight is the product of the components,

$$W_{ak} = BW_a \cdot WCF_{ak} \cdot NI_a$$

BW_a = the base weight (CATI or the Control as appropriate) for the sample area.

WCF_{ak} = special weight for the k-th person in the a-th area.

NI_a = single cell noninterview adjustment factor for the a-th area.

δ_{ak} = indicator function for characteristic of interest (i.e., in cell (i,j))

= 1 if in cell

= 0 otherwise

Note that the noninterview adjustment factors for CATI and the Control will be computed the same way but the actual values of the factor (NI_a) may differ for the two procedures.

4. Variance Estimation

We will use a stratified jack-knife technique of variance estimation. We will use this technique, not only to estimate variances on levels and rates of labor force data, but also in the log linear analyses performed by CPLX. We expect this procedure to produce reasonably accurate variance estimates. An advantage of this procedure is that the variance estimates for the cumulative averages automatically will account for the correlations associated with interviewing identical units over several months.

B. Response/Nonresponse Rates

1. Noninterview

a. General

The interview and refusal rates will be analyzed monthly. Comparison of the interview rates and refusal rates will be made between the full CATI sample and the Control. At this time we do not plan to perform hypothesis testing on noninterview rates. However, we shall have NI data tabulated by random group and carried in a file. Thus, we will have the ability to estimate variances on NI data if this proves desirable. For the most part we intend to produce descriptive reports and graphs of these data.

b. Key Variables, Tabulations and File Needed for the Analysis

Key variables will include the following:

- interview/noninterview status
- type of noninterview
- interview methodology

These analyses will use the CATI Analysis File. This file will be produced by the CPS Programming Branch in DSD. It will be the result of merging the Bridge Facility file, the CATI Field file, and the CPS MIS 1-4 sample in CATI Areas file. The CATI Analysis File will be used for generating all necessary interview/noninterview tabulations. This file also will be the source of the files used in comparing CPS and CATI labor force data.

For these analyses the overall CATI noninterview rates will be compared with the rates for the corresponding CPS sample. More detailed tabulations will be produced for operational and administrative purposes.

For administrative purposes, tabulations of interview/noninterview status by month-in-sample will be required. For interviewed households, additional analyses pertaining to type of interview (personal/telephone), will be performed. Outlines of these tabulations are displayed in Attachment B-1.

2. Item Nonresponse and Completion Rates

a. General

Completion rates will be used to compare the CATI and CPS methodologies for item nonresponse. The completion rate is the percent of entries made by interviewers which are complete and require no allocation while passing through the consistency edit. ("Blank" items are the complement of complete items). The completion rates for selected characteristics for the full CATI and the Control sample will be compared each month; however, we do not plan to conduct hypothesis testing for these data.

Blank items can be obtained differently according to the interview methodology used. Blanks are obtained during a traditional field interview for the following reasons: interviewer skip pattern error, respondent unable to provide requested information or respondent refuses to provide requested information. (The current CPS questionnaire does not permit "don't know" or "refusal" responses to be recorded.) During a CATI interview, the equivalent of blanks occur only when a "don't know" or "refusal" response is obtained. The CATI instrument does permit entries of "don't know" or "refuse." (These entries are recoded to blanks in order to pass them through the CPS processing system.) The CATI methodology eliminates interviewer skip pattern errors. The CATI advantage achieved by eliminating skip pattern errors can be measured by comparing completion rates for the two methodologies. Because of this advantage, we might expect CATI to have the higher completion rate.

However, there are factors which may make the CATI completion rate lower than that for the Control. Because CATI permits the interviewer to record refusals and "don't know" (DK) responses, the CATI interviewer may use this option, rather than probing a little harder to obtain a valid response (which is what the CPS interviewers must do). Another factor which acts to lower the CATI completion rate is that the CPS interviewers are able and encouraged to make call-backs to obtain missing data (DKs and refusals), but the CATI interviewers are unable to do this. The CATI case management software does not allow call-backs for missing data. It is conceivable that such a feature could be incorporated into the CATI system. This study will allow us to measure whether CATI's effectiveness in preventing skip pattern errors outweighs its weaknesses with regard to DK and refused items.

The CPS questionnaire will be redesigned in November 1986 to include refusal circles for Items 25C and 25D (hourly and weekly earnings). This means we will be able to make direct comparisons of refusals between CATI and the Control for these items.

b. Key variables

Key variables include:

Item 20A completion rate
 Item 20C completion rate
 Item 22A completion rate
 Item 22E completion rate
 Item 25C completion rate
 Item 25D completion rate

Item 25C refusal rate
 Item 25D refusal rate

The CATI Analysis file will be used for this analysis. The proposed tables are outlined in Attachment B-2.

c. Formulas

The completion rates will be computed for each item as

$$\text{Rate} = A/U$$

A = number of cases with item answered
 U = number of cases requiring an answer

The number U represents the universe for the completion rate and will be obtained by counting the number of cases where the edited field for the item contains a value. We assume here that the set of cases requiring a response to the item corresponds to the set containing a value after the edit.

The number A is the count of cases containing a value in the unedited field for the item.

The refusal rate will be computed for Items 25C and 25D as

$$\text{Rate} = R/U$$

where U is defined as above.

R = number of cases with item refused.

Here R is the count of cases in which the unedited field for the item is coded as "refused".

C. Response Variance

The most common method of measuring response error relies on reinterviewing sample units and comparing the original and reinterview responses. A less desirable, but still useful alternative is to compare month-to-month gross changes between labor force categories for the CATI and CPS methodologies. The methodology with more gross change is considered to be subject to greater response variability.

1. Hypotheses

We will test the hypothesis that the distribution of gross changes in labor force status is the same for the two methodologies. A difference in distributions indicates that response variance affects the methodologies differently.

2. Key variables, Tabulations and Files Needed for Analysis

We will obtain estimates of labor force status for the total population in identical rotation groups for two consecutive months. These will be cross-tabulated for both interview methodologies, as illustrated below. Using the data in these tables we will perform a log linear analysis to compare these distributions of gross changes, similar to the analysis described in II.A. above. Only records for persons interviewed in both months will be included in this analysis.

In this analysis Month 1 will use months-in-sample (MIS) 1, 2, 3 and Month 2 will use MIS 2, 3, and 4. This analysis will require special data files containing random group estimates of each gross change category in the example below.

EXAMPLE OF GROSS FLOW TABULATION

Labor Force Status, Month 1

Labor Force Status Month 2	Employed	Unemployed	Not in Labor Force
Employed CATI CPS			
Unemployed CATI CPS			
Not in L.F. CATI CPS			

3. Methodology, Formulas, Models

As mentioned, we will use log linear analysis to compare the distributions of the populations over the nine gross change cells in the table above for CATI and the regular CPS. This analysis is based on the same models used in the comparisons of labor force data. This analysis will tell us whether gross flows are different between CATI and the Control, but not which methodology experiences higher gross changes. Another statistic will indicate which methodology has greater gross changes, the percent of diagonal in the gross flow table. The methodology with a greater proportion off the diagonal (i.e. changed) experiences higher gross changes. From this we infer that the methodology experiences greater response variation.

D. Coverage

1. Hypotheses

We want to determine whether coverage achieved under the CATI methodology equals that under the regular CPS. To this end we will test the hypothesis that the distribution of persons by age, sex, and race in CATI equals that in the Control. Any significant difference in distributions will be taken to indicate that coverage differs between the two methodologies by demographic category. We will also compare within-household coverage, testing the hypothesis that the average number of persons 16+ per household is the same for the two methodologies. We do not plan to measure the more detailed aspects of within-household coverage, such as the distributions of secondary families, subfamilies, and unrelated individuals. We will compare CATI and Control population estimates with the corresponding "independent" estimates, which we plan to obtain from Population Division. These independent estimates will help us decide which methodology suffers least from problems of under- and over-coverage. We will be particularly interested in the coverage of those demographic groups less likely to have telephones. The files requested for the labor force analyses will provide the data we need for the coverage analyses as well. The persons per household data will be carried on a data file with selected noninterview data, and the age/race/sex population estimates will be included on the demographic by labor force file.

2. Tabulations

We will produce tabulations of the population 16+ in the demographic categories listed in section II.A. for CATI, the regular CPS, and for convenience, the ratio of CATI to the regular CPS. Finally we will tabulate the number of persons 16+ years old per household. Attachment D-1 shows the desired layout of these tabulations.

3. Models

We will use log linear analysis to test whether the demographic distributions of the population estimated by CATI and the Control differ.

We plan to display "pseudo-coverage ratios," the ratio of the CATI population estimate for each demographic category to the corresponding Control population estimate. These ratios will help to pinpoint the demographic groups experiencing the greatest differences in coverage between the two procedures.

We will compare estimates of the average number of persons per household using the t-test. This average will be computed as,

$$P = P / H$$

where

$$\hat{p} = \sum_{a=1}^{30} \sum_{k=1}^{h_a} W_{ak} \cdot P_{ak} \quad (4)$$

is the estimate number of persons in the study area

$$\hat{H} = \sum_{a=1}^{30} \sum_{k=1}^{h_a} W_{ak} \quad (5)$$

is the number of households in the study area, and

W_{ak} = the final weight for the household. Its components are defined similarly to the corresponding terms in equation (3).

$$W_{ak} = (BW_a)(NI_a)(WCF_{ak})$$

WCF_{ak} = weighting control factor for the household.

P_{ak} = the number of persons in the k-th interviewed household in the a-th sample area.

Attachment D-2 shows the layout of this table. At this time we do not plan to examine CPS and CATI household sizes by race or ethnicity. We considered comparing the two methodologies for coverage of subfamilies, secondary families, and unrelated individuals but decided this additional information would not justify the expenditure of resources required.

E. Characteristics of Respondents

These analyses have been given extremely low priorities. Given the current levels of available resources, it is unlikely that these analyses will be performed. At this time, none of the files needed for this research have been requested.

1. Demographic Distribution of Respondents

We could tabulate the demographic characteristics of the respondents only (rather than respondent and all eligible household members) interviewed by CATI and the Control using the demographic categories described in II.D. We would use log linear analysis to determine whether the distribution of respondents by demographic category differs between CATI and the Control (See Attachment D-1.).

The hypotheses tested, tabulations and data files, and models and formulas used will be similar to those described for the coverage comparison, but for respondents only, rather than for all sample persons.

2. Identical Respondents from Month-to-Month.

In any two consecutive months a household may have the same respondent or two different respondents. A difference in the proportion of identical respondents might result in differences in gross flows or otherwise affect data quality. We will use the t-test to examine the hypothesis that the proportion of identical respondents is the same for CATI and the CPS.

We will examine tabulations showing identical and different respondents for each pair of consecutive months. We also will accumulate these data as the study continues. Summary data files of these tabulations will be obtained.

To make this comparison we will estimate the number of identical respondents from month-to-month, and the number of total respondents in households interviewed in

both months (\hat{R}). Again we will estimate variances via the stratified jack-knife technique.

$$\hat{S} = \sum_{a=1}^{30} \sum_k^h (W_{ak}) (I_{ak}) \quad (6)$$

= identical respondents.

$$\hat{R} = \sum_{a=1}^{30} \sum_k^h (W_{ak}) \quad (7)$$

where

h_a = the number of sample households in the i -th area that were interviewed in both months.

I_{ak} = indicator for respondent at the a, k -th household,

= 1 if respondent is identical

= 0 otherwise

$W_{ak} = (BW_a)(WCF_{ak})$ = final weight of the a, k -ith household without the noninterview adjustment, with BW_a and

WCF_{ak} defined as in equation (5).

F. Quality of Industry and Occupation data.

1. Cases Requiring Referral During Industry/Occupation Coding in Jeffersonville

a. General

Comparisons of industry/occupation referral rates will be made between the full CATI sample and the CPS control sample.

b. Key Variables, Tabulations, and Files Needed for the Analysis

Key variables will include the following:

- Major industry classification
- Major occupation classification
- Class of Worker classification
- Industry Referral Rate
- Occupation Referral Rate

For this analysis we need the CATI Analysis File.

NOTE: Due to budget constraints, CATI industry/occupation data will be coded for only three months, so comparisons can be made for only those months. If funding becomes available later, we can reimplement the coding procedures at any time.

c. Methodology, Formulas, and Models

This analysis will not include any hypothesis testing. A report will be issued covering the period in which the data are available.

2. Distributions of Industry and Occupation Classifications

We will compare the CATI and CPS distribution of industry and occupation classifications for the major level categories (listed below). We will use weighted tabulations of unedited codes to determine how many blank and uncodable responses occur under both procedures. We again will rely on log linear analysis to test for differences between the distributions. This will not allow us to decide that one methodology is better than the other, merely that there are differences. Because current plans call for industry and occupation coding to be performed only for November 1986 through January 1987, these data will not be as reliable as the labor force data. Finally we will compare the proportions of DK/NA responses obtained from the two methodologies.

Industry Categories

Manufacturing
Wholesale and Retail Trade
Finance and Services
Other
DK / NA

Occupation Categories

Managers and Professionals
Technical, Sales, and Administrative Support
Operators, Fabricators, and Laborers
Service and Other occupations
DK / NA

G. Telephone Availability

It would be interesting to carry out many of the above analyses using only those cases with no telephone available. The primary purpose would be to determine whether CATI cases interviewed in the field are treated differently from similar Control cases. We know that all CATI "telephone not available" cases are interviewed in the field. All things being equal, the corresponding Control cases should produce similar estimates and data quality measures. If CATI field cases produce different estimates, this introduces a confounding effect into the study, which must be considered when the data are analyzed and conclusions are reached.

Unfortunately, because only about 8 percent of all households have no telephone available, ^{1/} we can expect only about 200 such cases per month in the CATI sample. Telephone availability is a control card item which currently is transcribed to the CPS questionnaire (CPS-1) only in March, July, and November. Due to cost constraints, this item will not be transcribed each month, which reduces the sample available for analysis to such a level that there is little point in performing most of these analyses for no telephone households.

The only such analysis likely to justify its cost is a comparison between CATI and the Control of the proportion of sample cases with no telephone available. This comparison would be of value primarily to determine whether the interviewers for some reason are recording telephone availability differently for the CATI and regular Control samples. Since telephone availability determines which cases are assigned to the CATI facility, it is important to assure there are no differences in determining telephone availability. It is worth mentioning that we do not expect to find any such differences. We will make this comparison using the t-test, estimating the proportion of non-telephone households by,

$$\hat{A} = \hat{T} / \hat{H}$$

where

$$\hat{T} = \sum_a^{30} \frac{h_a}{\sum_k} W_{ak} T_{ak} \quad (8)$$

$$\hat{H} = \sum_i^{30} \frac{h_i}{\sum_j} W_{ak} \quad (9)$$

and

t_{ak} is the indicator for telephone available in the (a,k)th interview sample unit.

= 1 -- telephone not available

= 0 -- otherwise.

h_a and W_{ak} are as defined for equation (5), including the NI adjustment factor.

^{1/} Based on CPS telephone availability data.

We will produce tabulations of these cases. Approximate variances will be used for the any hypothesis testing.

III. SCHEDULE AND PRIORITIES

This section lists the activities required to carry out the analyses discussed in this plan. It also lists priorities for the various analyses, in the event that lack of resources does not permit us to perform all the analyses.

A. Schedule

<u>Activity</u>	<u>Approximate Target Date</u>
1. Write specifications to modify CPS data file and create files needed for analyses.	09/15/86
2. Write software to create the files needed.	11/15/86
3. Write software to produce the tabulations needed.	12/15/86
4. Write software to estimate variances.	12/15/86
5. Produce files needed.	12/22/86 and monthly
6. Produce tabulations and variance estimates.	12/30/86 and monthly
7. Prepare Preliminary Report -- findings based on 5 to 6 months of data.	05/01/87
8. Prepare Final Recommendation	12/01/87
9. Prepare Final Report -- complete documentation of methodology of the study, detailed findings, quality of data, etc.	01/31/88

B. Analysis Priorities

The analyses described in section II. are grouped here in order of priority. Priority was determined by the expected utility of the information we plan to obtain from the analyses. The analyses have been listed in five priority groups, with the first group having the most priority. Projects in group V have a very low probability of completion.

<u>Group</u>	<u>Analysis Project</u>
I	Response Rates
I	Labor Force Comparisons
I	Coverage Comparisons -- Demographic groups
II	Response Variance -- gross flows
II	Industry and Occupation coding -- distribution
III	Coverage -- Persons per household
IV	Percent of units with telephone available
IV	Identical respondents
V	Characteristics of Respondents
V	Labor Force estimates by Month-in-Sample

SMD/JBushery/dlc/lgs/vw-2747/JBU30

Attachment A

Labor Force Estimates For CATI Phase II

Replicate by current month and cumulative months

TOTAL FOR 30 AREAS	Total Population in Demographic Group	CLF	CLF RATE	Employed	Unemployed	Unem- ployment Rate	Not In Labor Force
--------------------	---	-----	-------------	----------	------------	---------------------------	--------------------------

CATI

Replicate by Age: Total 16+, 16 - 19, 20 - 24, 25+
by Race: Total, Black, White, Other, Hispanic
by Sex: Total, Male, Female

CPS

Attachment B-2.

CATI Phase II
Accuracy Rates
for Selected Items

Replicates by current month of interview
and cumulative months

CPS CATI

Total Accuracy Rate (All items)

Accuracy Rates for Selected Items

Item

20A

20C

22E

25C

25D

Refusal Rate

25C

25D

Attachment D-1.

Population Estimates by Demographic Group
For CATI Phase II Analysis

Replicate by current month of interview and cumulative months

Methodology -----	Demographic Category (Race, Sex, Age) and Demographic Cell Number								
	...White Male....			...White Female....			...Other Female....		
	16-19	20-24	25+	16-19	20-24	25+	16-19	20-24	25+

"Pseudo" coverage ratios also will be computed for these demographic categories.

CPS

CATI

Attachment D-2.

Tabulations For CATI Phase II Coverage Comparisons
Persons 16+ Per Household

Replicate by current month and cumulative months

Sample Area	Regular CPS	CATI	Difference CPS minus CATI
Total for 30 Sample Areas			
Number of Households			
Number of Persons 16+			
Persons per Household			

Attachment F-1.

CATI Phase II
Referral Rates for
Industry and Occupation Entries

CPS

CATI

Total Records Coded for Industry
Total Industry Referrals
percent referred

Total Records Coded for Occupation
Total Occupation Referrals
percent referred

Attachment F-2.

CATI Phase II
Comparison of Industry
and Occupation Estimates

Industry Categories

Persons in Category
CPS CATI

Manufacturing
Wholesale & Retail Trade
Finance and Services
Other
Don't know/NA

Occupation Categories

Managers & Professionals
Technical, Sales, & Administrative Support
Operators, Fabricators, and Laborers
Service & Other occupations
Don't know/NA