

Collecting Census Long Form Data Over the Telephone: Operational Results of the 1995 CM CATI Test

Kenneth B. Dawson, Janice A. Sebold, Susan P. Love, Lynn Weidman

U.S. Census Bureau
Washington, DC 20233

The paper was presented at the Annual Meeting of the American Statistical Association (ASA), August 1995.

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

I. BACKGROUND

Every 10 years the Census Bureau embarks on the enormous task of counting every person in the United States. In addition to collecting population and housing counts and basic demographic information, the decennial census obtains detailed data from a sample of households. This sample information, often referred to as long form data, is crucial to Federal, state, and local governments for planning and distribution of governmental funds. Private companies also make major business decisions on programs and day-to-day operations based on decennial census data. Considering the many users of these data and the constantly changing face of the Nation, the Census Bureau is responding to the ever increasing need for more frequent collection of the sample data. The Continuous Measurement Survey (CMS) is the vehicle the Census Bureau is proposing to use to collect and distribute this information on a more timely basis over the decade.

The CMS will be a large monthly household survey implemented in 1999. It will use a tri-modal data collection methodology which will 1) Mail or deliver to a large sample of addresses, 2) Attempt to interview the nonrespondents by phone, and 3) Select a subsample of the remaining nonresponding sample units, upon completion of the telephone follow-up operation, for personal interviews.

The mail component of the test relies on self-response using a paper questionnaire, while the telephone and personal interview follow-up will be computer assisted. We will begin testing all three data collection modes and their integration in 1996.

Beginning in November 1994, we tested the computer assisted telephone interviewing (CATI) phase through the 1995 Continuous Measurement CATI Test. This test had two objectives. The first objective was to test our ability to gather long form data over the telephone. To help with this evaluation, we tested different wording and reference periods. The second objective was to gather operational and cost data to help with planning of the 1996 test and subsequent phases of the CMS.

The 1995 test was the first Census Bureau survey to use the CASES authoring software and the new CENCATI Control System in production. November data were used only to ensure that the system was working properly. All analyses in this paper use data for December 1994 through April 1995.

The purpose of this paper is to discuss selected results of the 1995 test. Information on the sample, vendor-supplied telephone lists, response and refusal rates, length of interview, best time to call, and the number of calls to complete a case are presented in this paper.

II. SURVEY DESIGN AND SAMPLE SELECTION

The 1995 test was a two panel "cold-contact" CATI survey. Respondents at selected telephone numbers received no advance notification before being called. The differences between the two panel instruments were question wording and reference periods. The test was conducted from the newly established census telephone center in Jeffersonville, Indiana.

The monthly samples of telephone numbers were drawn from commercial databases of U.S. telephone listings on CD-ROM. The listings that the Census Bureau purchased for the 1995 test were distributed by the software vendors PhoneDisc and Select Phone. For the first three months, 4,000 telephone numbers were selected each month. In March and April the sample size was reduced to 2,000 because the two panels were combined into a single instrument.

Three stages of sample selection were used to obtain complete telephone numbers. 1) Area code (AC) selection (three digits). 2) Prefix or central office code (three digits). 3) Suffix selection (four digits).

The monthly samples were selected as follows:

1. Each month 40 three-digit area codes were systematically selected from the list of valid area codes. Specialized numbers such as 800 numbers and commercial area codes were excluded from the list before selecting the sample. Upon establishing a list of eligible area codes, a sampling interval (TE) and initial area code (SW) were selected. The TE was determined by dividing the number of eligible area codes (N) by the desired number of area codes (n) for the month.

To select the SW, a random number (RN) between 0 and 1 was calculated and then multiplied by the TE. The result of this equation was rounded to the nearest integer and called SW.

2. The SW and TE numbers were applied to the list of eligible area codes. First the SWth area code on the list was selected then each TEth ordered

- area code for the rest of the list was selected.
3. After selecting a list of area codes, a three-digit number was randomly generated for each selected area code. This three-digit number became the telephone number prefix for the area code.
 4. Upon establishing a list of 40 six-digit AC/Prefix combinations, the combinations were manually matched to PhoneDisc and Select Phone software databases. When a match was obtained, the AC/Prefix was selected. If a prefix was not found on the list for a particular area code, the next prefix in sequence was selected. A list of selected AC/Prefixes was maintained to eliminate the possibility of re-selection during sampling for subsequent months. Area codes could be selected for multiple months' samples but no AC/Prefix could be selected more than once.
 5. Once an AC/Prefix was selected, the first 100 residential listings were selected for interview. The selected listings were downloaded to ASCII text files. The information downloaded contained the address and telephone number for the selected households. This information was the basis for the CATI input files.
 6. The selected listings on the file then were alternately split between the two panels to ensure comparability between the panels.

III. SAMPLE LISTING SOURCE

Samples for the 1995 test were selected using commercially available telephone listings on CD-ROM. These lists were beneficial for a number of reasons. The lists were easy to obtain through the mail, easy to use for a small sample, and inexpensive, at less than \$200 for quarterly update. The sources used for the test were PhoneDisc and Select Phone.

Since the main focus of the 1995 test was to evaluate the CATI instrument wording and procedures, telephone lists on CD-ROM were the best solution. These samples cannot be weighted to obtain a national estimate because of the method of selection. They were evenly distributed between the two panels; that is, telephone numbers from each area selected were alternately split between the panels of the instrument. The result was that characteristics of persons living at selected telephone numbers varied from month to month but were similar among the panels within a given month.

In 1995, the ultimate sample unit (USU) was telephone number. If a selected number worked, the interviewer confirmed that the correct telephone number had been reached. The interviewer then read the address as found on the listing source and asked if the address was correct. If not, the correct address was requested and captured in the instrument. [Table 1](#) (see end of text for tables) is a summary of results by month of the survey for the address verification.

The results in [Table 1](#) show that the PhoneDisc source had a higher percentage of correct addresses contacted than Select Phone. Nearly 23 percent of the January sample telephone numbers and 27 percent of the February sample, which were both selected from Select Phone, did not have address as part of the record. In fact, street address was not a required variable on the Select Phone databases. These listings were excluded from these address analyses. PhoneDisc has a street address for all listings.

In addition, PhoneDisc also had more correct numbers. Between 58 percent and 64 percent of the PhoneDisc numbers had the correct address associated with them. Overall, PhoneDisc had a correct address for 60 percent of the verified telephone numbers. While Select Phone had a correct address for only about 50 percent overall.

A fully operational CMS will rely heavily on our ability to find telephone numbers for addresses without the benefit of surnames. PhoneDisc and Select Phone provided us with inexpensive and easy to use sample sources. However, their failure to provide a higher match rate between telephone number and address, combined with the large number of disconnected telephone numbers, renders both sources as unacceptable as the main source for telephone numbers for CMS.

V. RESPONSE AND REFUSAL RATES

The response rate for the 1995 CM survey was 65.2 percent (see [Table 2](#)). The average refusal rate was 21.1 percent. These rates are based only on the total number of eligible telephone numbers, which included all working numbers that were connected to residential housing units. Telephone numbers which did not reach a private residence were considered ineligible for the survey. Ineligibles included mobile or pay phones, businesses, special places such as college dorms, FAX numbers, and disconnected telephone numbers.

Since the focus of the 1995 test was not response and refusal rates, no extra measures were taken to improve these rates. However, interviewers did attempt to call all refusals a second time and were able to convert an average of 25 percent each month. The main reason that respondents gave for refusing to cooperate with the interviewers for the first few months was the expected length of the interview. Respondents were told that the average interview would take 45 minutes. After several months, the average time to complete an interview was found to be just under 30 minutes and the introduction was changed in February to reflect this estimate. After this change, the main reason given by respondents for refusing went from the expected length of interview to that they just did not want to participate.

The rates for the 1995 test are comparable to other cold contact surveys conducted by the Census Bureau, such as the Consumer Point of Purchase (CPP) Survey with response rates as high as 66.1 percent and refusal rates of 13.9 percent. Since subsequent CMS phases will use CATI as a follow-up for the nonrespondents of the mailout, these response and refusal rates will not apply.

VI. SCHEDULING EFFICIENCY

Scheduling an efficient CATI staff is crucial to keeping costs low, productivity high, and data quality good. When planning an efficient CATI survey, the main questions that must be answered are "What are the best times to call respondents?" and "How many times should a household be called?". The latter issue is especially important to the CMS because of the tri-modal design. Some cutoff or time limit must be established so that CATI nonrespondents can be sent to the computer assisted personal interviewing (CAPI) phase on a timely basis.

To identify the optimal times to reach the largest number of respondents, the efficiency throughout each month and day was analyzed. The number of individual calls by outcome were compared for different time periods throughout the day and by days of the week. To simplify analysis for the 1995 test, outcomes were grouped by the following five general categories: 1) Interviews - fully completed interviews for each roster person in the household. 2) Unusuables - includes partials with no callback set, household deceased, mobile phone, pay phone, business phone, college housing, no one uses place as usual residence, FAX number, unknown residential status, unavailable through closeout, unconvertible language barrier, disconnected or non-working number. 3) Refusals - includes hostile breakoffs, persons who refused to state residential status, or immediate hang-up. 4) Callbacks - includes partials with callback set, language barrier (referred to supervisor), or answering machine where number was definitely reached. 5) No Answers - includes unanswered calls, normal busy or circuits busy, fast or WATS busy, number could not be completed as dialed, no signal, funny signal, bad connection, wrong number dialed or reached, possible wrong number, answering machine-unknown if reached correct number, or call not attempted now.

The hours of the day were grouped into morning, afternoon, and evening hours. Morning included all calls made before noon; afternoon included all calls made from noon until 4:59 p.m.; and evening calls were made from 5:00 p.m. on. All time analyses are based on the respondent's current time, taking time zone into consideration. Time zone is important to staffing a telephone unit because interviewers must work during the times when respondents are available. The days of the week were examined individually and were also grouped into two categories, weekdays and weekends. Weekdays included calls made Monday through Friday. Weekends were Saturdays and Sundays.

For certain analyses we categorized the call outcomes by phase. Phase 1 is the period during which the sample telephone numbers were initially being attempted. The beginning of Phase 1 is the first day of calling for a particular month. The end of Phase 1 is the day in which all or most cases have been called at least once. Phase 2 is the remaining days of the month after Phase 1 is completed.

Using these groupings, outcomes per call by time of day were compared to the active hours throughout the day. Active hours include all time spent on a case, including the time necessary to place the call and any time spent talking to respondents. The active time counter started when a case was retrieved and stopped after the interviewer exited the case.

It was determined that the most productive hours of the day are the evening hours. In addition, productivity for all time periods increased during the weekends. Since evenings and weekends are times when most working people are home and are available, the number of interviews completed compared to the amount of time spent making calls was higher during these periods.

These findings have a direct impact on staffing for the telephone center. The bulk of interviewers for the survey should work during the times that most respondents are at home. The number of interviewers should be lower during the least productive periods of the day. If too many interviewers are staffed during less productive periods then interviewers will spend more time making unproductive calls. \

A. Time of Day

As a measure of efficiency, call outcomes were compared to the number of active hours during several periods of each day throughout the survey. [Table 3](#) shows the total number of calls and the percent of calls by outcome for each of the three time period groupings.

Overall, the least productive part of each day included both the morning and afternoon. These periods made up nearly three quarters of the active hours but in absolute numbers terms had only 65.7 percent of the interviews. While the absolute number of interviews was higher during the day, the percent of calls spent on interviews nearly doubled during the evening hours. Calls to interviews during the morning and afternoon comprised 7.3 and 7.5 percent, respectively, of all calls during each time period. This proportion nearly doubled to 13.8 percent of the calls during the evening. In addition, the ratio of complete interviews per active hour was lower during the day than in the evening hours. The number of interviews per active hour during the day was 1 for morning and 1.1 in the afternoon. The ratio of interviews to active hour was 1.4 during the evening hours which is 40 percent increase over the morning and a 27% increase over the afternoon. These are significant difference. Because the ratio of interviews per active hour was low during the day, interviewers made more unproductive calls. During the day interviews had about seven unanswered calls per hour. The number of unanswered calls per active hour was reduced by 44 percent to 3.9 during evenings. About half of all calls during the morning and afternoon went unanswered.

The most productive period of the day throughout the survey was after 5:00 p.m. While only 27.5 percent of the active hours occurred during the evening, 34 percent of the interviews were obtained. Given these patterns, we feel efficiency will rise if more interviewers are staffed during the peak hours of the day, between 5:00 p.m. and 10:00 p.m. During the 1995 test, the bulk of the active hours occurred during the least productive period of the day.

B. Day of Week

Traditionally, Census Bureau telephone units have been staffed during the daytime on weekdays and Sunday. Partly because Friday evenings and Saturday evenings had been considered less productive times to conduct CATI interviewing. During the 1995 test, however, we found that weekday evenings and weekends in general were the most productive times. We found that these times were the most productive periods when looking at the number of interviews to the active hours (see [Table 5](#)). Table 5 shows that the productivity for any evening hour was higher than any morning or afternoon period with the exception of Saturday morning and afternoon which were comparable to weekday evenings at 1.4 interviews per active hour. The evening hours were consistently more productive than the morning and afternoon periods.

VII. DETERMINING CUTOFFS

Because of the tri-modal design of the Continuous Measurement Survey, we must determine a "cutoff" point for cases that have not been

interviewed. Determining appropriate criteria for sending cases to CAPI sampling is crucial to the smooth flow of a Continuous Measurement system. Cutoffs will be based on a combination of criteria, including how long the case has been in the CATI unit, how many calls have been attempted on the case, and the pattern of call outcomes observed for these calls.

Values and limits for these criteria must be determined, after which cases meeting any of these criteria limits would be removed from the CATI universe and would become a member of the nonresponse universe from which the CAPI subsample will be selected.

A. Time limits

In the 1995 CATI Test, the most productive period of each month was during Phase 1, the period during which the sample telephone numbers were initially being attempted. During the 1995 test, the number of interviews per active hour for all cases during Phase 1 was 1.4 (see [Table 4](#)) which is comparable to the 1.4 interviews per active hour achieved during the evening hours of the combined phases.

The comparison between phases within an interviewing period is an important one. Efficiency dropped dramatically during Phase 2. The total number of callbacks per active hour increased from 2.9 in Phase 1 to 5.4 in Phase 2. In addition, unanswered calls went from 3.5 in Phase 1 to 8.5 in Phase 2. Productivity decreased as the number of interviews per active hour dropped from 1.4 in Phase 1 to 0.9 in Phase 2.

This trend shows that the longer a case sits in the telephone unit the less likely it is to become a completed interview.

B. Limits on the number of calls for a case

To help determine the limits for the number of calls allowed a case, it was necessary to look at the outcomes of cases compared to the actual number of calls it took to obtain the outcome. This evaluation was helpful in demonstrating the overall efficiency of the calls.

When looking at interviews by the number of calls it took to complete the interview, it was found that 91.8 percent (see [Table 6](#)) of all interviews were obtained in ten calls or less. Over three quarters (78.4 percent) of the interviews occurred in five or fewer calls. While 8.2 percent of all completions did occur after the tenth call, they were mixed among much unproductive calling. Of the 28,500 calls placed after the tenth call made on a case (see [Table 7](#)), only 2.4 percent (693) of these resulted in a completion. It took 20 or more calls to interview 145 of these cases. Of the 28,500 calls made after ten calls, unanswered calls made up 62.1 percent (17,689) of these calls. Refusals accounted for 31.7 percent (9,043) of the calls. Combined, these two unproductive outcomes comprised 93.8 percent of the calls made after the tenth call to a case.

C. Determining cutoffs based on call patterns

We may be able to predict the likelihood of completing a case by looking at individual call outcome patterns. Patterns may consist of the number of callbacks or the total number of calls to a case. There is evidence that potential respondents may not refuse outright but may string along interviewers in an attempt to "avoid" the interview by repeatedly requesting callbacks. During the 1995 test only 10 percent of the interviewed cases had four or more callbacks prior to completing an interview. While 22 percent of the refusals had four or more callbacks before the final refusal, 62.5 percent of the refusals occurred on four or more calls. This pattern indicates that people may be reluctant to give an outright refusal and that they may attempt to avoid the interview by setting a callback. As shown in [Table 7](#), the proportion of callbacks and unanswered calls increases as the number of calls to a particular telephone number increases. 62.1 percent of the calls made after the tenth call were unanswered, 31.7 percent were callbacks. This increase is matched by the decrease in the interviews obtained. By the eleventh or more call only 2.4 percent were to complete an interview. With each additional call the likelihood of an interview diminishes.

IX. CONCLUSION

Much of what was learned in the 1995 CM CATI Test is being used in planning for the CATI segment of the 1996 test. The following summarizes what we have learned through the 1995 test and some of the conclusions that have been drawn from this information.

A. Vendor-supplied telephone lists

Although PhoneDisc provided a better match between telephone number and address than Select Phone, neither source was accurate enough to be used as the main source for telephone numbers for future CM matching. The sources on CD-ROM are also not accessible enough to gather telephone numbers for a large number of addresses. Based on past census experience, we will use a vendor or group of vendors such as Telematch who offer tape-to-tape services enabling us to get large quantities of telephone numbers inexpensively. We will also use an on-line telephone listing service as a secondary source when a number obtained through Telematch is not correct. PhoneDisc may be used as a backup to an on-line service.

B. Response and refusal rates

Although we were able to achieve a response rate of 65.2 percent in 1995, the CATI sample for 1996 will consist of mail nonrespondents. These persons will have already refused the mail component of the test.

We expect that mail nonrespondents may not be as cooperative as those selected in 1995. A significant portion of the interviews for 1995 would probably have been mail respondents, so our response rates in the 1996 test are likely to decline somewhat. We plan to work with the telephone

facility staff to alter staffing according to the results we found in 1995 in an attempt to sustain these rates through efficient staffing. We will also work with the telephone staff to develop a workshop and job aids for handling reluctant respondents in 1996.

C. Best time to call

We will recommend that the bulk of the CATI facility schedule staff to work between 5:00 p.m. and 10:00 p.m. on weekdays and during the afternoon on Saturday and Sundays. This is especially important for Friday and Saturday evenings because traditionally fewer interviewers have been staffed at these times, although, they have proven to be very productive time periods.

D. Limits on calls

We will limit the number of calls that can be made for a particular case to ten. The time a case stays in the CATI unit will be limited and limits will be placed as appropriate from our pattern analysis.