CATI and Touchtone Self-Response Applications for Establishment Surveys

George Werking, Alan Tupek and Richard Clayton[1]

Abstract: Computer-assisted collection methods offer the implementation of nontraditional approaches to the control of response and nonresponse errors in establishment surveys. Recent research for an establishment survey of 300 000 monthly reporters has focused on developing (1) an ongoing record check program using computer-assisted telephone interviewing (CATI) for controlling response error and (2) a fully automated touchtone data collection system for employers to quickly and easily self-report their data.

Key words: Nonsampling error; data editing; CATI; self-response.

1. Introduction

This paper addresses two distinct computer-assisted methods for controlling nonsampling error in establishment surveys.

Survey literature has traditionally focused on household surveys. More recently, however, attention has turned to errors in establishment surveys, specifically, the control and measurement of errors (OMB Statistical Policy Working Paper 15 (1988)).

Establishment and household surveys are similar in many ways and often have many of the same problems. However, the differences provide challenges for which new and creative approaches need to be developed. Establishment surveys suffer primarily from errors stemming from the recordkeeping system, while household surveys are often subject to

errors from proxy responses and recall, telescoping and other memory errors. In establishment surveys, it is usually the respondent's job to be familiar with the recordkeeping system. Both household and establishment surveys are subject to errors related to a misunderstanding of questions.

One of the traditional methods used to identify response variance is through screens and edits. These are particularly effective in surveys where data are compared over different reporting periods. However, these checks are usually too broad to detect many definitional inconsistencies and thus, insufficient for identifying and correcting consistently reported errors. An ongoing CATI record check program identifies the major sources of such errors in a large U.S. Bureau of Labor Statistics (BLS) establishment survey and provides the mechanism for correcting ongoing reporting.

[1] U.S. Bureau of Labor Statistics, Washington, D.C. 20212, U.S.A.

Other computer-assisted methods are used to improve response timeliness and to reduce attrition rates in low budget surveys. Research studies using CATI proved extremely effective when compared to current mail surveys (Werking, Clayton, Rosen, and Winter (1988)). However, the higher cost of CATI led to the development of a self-response system using touchtone recognition technology. The Touchtone Data Entry (TDE) system features digitized phrases which request answers from the respondent using the keys of a touchtone telephone. The responses are captured for subsequent editing and reconciliation. TDE offers the potential for reductions in survey collection costs by eliminating labor intensive activities such as mail handling and key entry. CATI is used to follow up the TDE nonresponse.

2.   Response Error Control

2.1.   Background

The Bureau of Labor Statistics (BLS) issues monthly employment estimates for the U.S.A. from two surveys: The Current Population Survey (CPS) and the Current Employment Statistics (CES) Program. The CES Program is a monthly establishment survey of 300 000 units conducted by BLS in cooperation with State Employment Security Agencies. The CES Program provides data on employment, hours, and earnings by detailed industry and geography.

The CES Program has less than three weeks to collect, key enter, edit, tabulate, and publish the preliminary national estimates, which are released concurrently with the employment and unemployment data from the household survey (CPS). The survey collection is primarily conducted by the local state agencies via mail. While this mode of data collection has cost advantages, it presents problems for controlling timely response for

monthly publication deadlines. Research tests conducted between 1984 and 1986, involving the replacement of mail collection with CATI, have shown CATI to be an effective means for increasing the survey response rates, lowering survey attrition rates and improving data quality.

Commonly used response error control techniques for mail collection rely heavily on the screening and editing of reported data. CATI collection in the CES Program allows on-line editing of reported data, including range checks, internal consistency checks, and longitudinal editing checks. This improves the customary editing and screening of mail questionnaires in that each respondent can be asked follow-up questions for any edit failure. However, even with CATI improvements, screen and edit control will generally catch only gross errors, and will offer little protection against widespread, systematic smaller errors, or large errors which are consistently reported in the data over time.

Current research demonstrates that a record check survey conducted with the respondents can identify and correct many of the systematic and consistent errors which traditional screens and edits miss. The CATI record check approach can be economically and operationally feasible for reducing response error in panel surveys. This extends the more traditional use of personal visit record check surveys for measuring error to ongoing control of response error by periodically rotating through the entire sample survey using a CATI record check approach.

2.2.   Response environment

Exhibit A compares the response environment for establishment surveys and household surveys. The examples provide key questions related to employment from the Current Employment Statistics Program, and the Current Population Survey, respectively. The examples illustrate some differences be-

Exhibit A

Response Environment for Questionnaire Data Items From Establishment Surveys and Household Surveys

	Establishment Surveys	Household Surveys
Example of a Data Item	● "Report the number of paid employees who worked or received pay for any part of the pay period which includes the 12th of the month."	● "Did you do any work at all last week, not counting work around the house?"
Source of Respondent Information	● Existing records, knowledge of recordkeeping system	● Knowledge of characteristic, activity, or event
Potential Retrieval Problems	● Access to records ● Timeliness of records ● Retention of records ● Definitional differences of records to data items ● Unfamiliarity with records ● Misunderstanding of data item	● Memory (recall, telescoping) ● Unfamiliarity with characteristics, activity or event ● Misunderstanding of data item
Respondent Characteristics	● Responsibility – identification of person whose job it is to be familiar with records ● Availability – standard work hours ● Stability – low job turnover ● Comprehension – less variable, since questions relate to respondent's job (payroll, personnel, accounting, etc.) ● Data supplied are less likely to have personal interest	● Respondent rules – for example: "knowledgeable adult" ● Availability – uncertain ● Stability – may be different household member ● Comprehension – more variable ● May have personal interest in data supplied
Questionnaire Objectives	● Use concepts & definitions customarily recognized by targeted respondents ● Response mechanism is to produce linkage to firm's record & respondent's knowledge of records	● Use questions easily understood by general population ● Response mechanism is to produce linkage to respondent's memory
Interviewer Objectives	● Educate respondent on the relation of their data to survey estimates to encourage adherence to survey definitions	● The relation of the respondent's answers to survey estimates may not be disclosed in order to avoid respondent conditioning

tween establishment and household surveys. The number of employees in the CES Program will likely be reported from records maintained by the payroll department of an establishment. This requires the respondent to be familiar with the recordkeeping system, while the household respondent needs to be knowledgeable of a specific characteristic, activity, or event.

The potential for response error exists in both household and establishment surveys, however, for different reasons. While household surveys are often subject to errors due to proxy responses and recall, telescoping and other memory errors, establishment surveys suffer primarily from errors stemming from the recordkeeping system. The ability of the respondent to access the necessary records, the timeliness and age of recorded data, respondents who are unfamiliar with records, as well as the definitional differences between the establishment's records and the survey are all sources of response (or nonresponse) error. Both household surveys and establishment surveys are subject to errors related to a misunderstanding of questions.

In establishment surveys, it is usually the respondent's job to be familiar with the recordkeeping system. These respondents are, in general, familiar with the concepts of the survey. For example, payroll, sales, or occupational data would usually be collected from payroll departments, accounting departments, and personnel offices, respectively. The respondents are also likely to be reporting data to other government surveys. In establishment surveys, concepts and questions are developed to be comprehensible to a specific type of respondent. In household surveys, concepts and questions must be designed to be understood by the general population, which will span a far more diverse range of age and education.

Questionnaires and interviewer procedures are also designed with different objectives in household and establishment surveys. In household surveys, respondents are often not informed about the classification system (i.e., employed, not employed, in labor force, not in labor force). This is done to prevent respondents from providing socially desirable answers. In establishment surveys, the respondents are not reporting data about themselves, or their families. Therefore, it is assumed that respondents are less likely to be motivated to provide socially desirable answers rather than correct answers.

2.3. Concepts

Response error is a major component of potential nonsampling error in surveys, which is both difficult to detect and expensive to measure. At aggregate levels, these types of errors have been shown to dominate the error structure for survey estimates. Major sources of response error include: poorly formulated concepts, definitions, and questionnaires; nonexistent or inaccurate records on the characteristics of interest; inadequate interviewer training; and unknowledgeable respondents.

Given the survey error models proposed by, for instance, Hansen, Hurwitz, and Bershad (1961):

$$y_{i\alpha} = \mu_i + \beta_i + d_{i\alpha},$$

where

μ_i is the true value for the ith unit,

β_i is the bias component for the ith unit, and

$d_{i\alpha}$ is an error component which follows some frequency distribution.

The mean square error (MSE) of the mean $\bar{y}$ of a simple random sample (SRS) of n elements, under the assumptions that:

$$\mathrm{E}\,(\beta_i) = 0,\ \mathrm{E}\,(\mu_i + \beta_i) = \mu$$

$$\mathrm{E}\,(d_{i\alpha}) = 0,\ \mathrm{E}\,(d_{i\alpha}\,d_{j\alpha}) = \varrho_R \sigma_R^2$$

is:

$$\mathrm{MSE}\,(\bar{y}) = 1/n\,(S_\mu^2 + \sigma_R^2\,[1 + (n-1)\varrho_R]) + B^2,$$

where

S_μ^2 is sampling variance,

ϱ is intraclass correlation,

σ_R^2 is simple response variance, and

B is bias.

This model and decomposition of the MSE of $\bar{y}$, and variations of this model, have been used for measureing nonsampling errors in household surveys. Little is written in the literature about nonsampling errors in establishment surveys. The simple random sampling (SRS) model is inappropriate for establishment surveys and more complex models would be difficult to develop and analyze. However, one can first consider a simple case of an establishment survey collected by mail and selected by SRS, where no special emphasis is placed on larger establishments. In this case, one might consider ignoring the intraclass correlation between coders and analysts. The intraclass correlation in household surveys, primarily due to interviewer variability, can be expected to be much larger than variability of estimates due to coder and analyst variability. One would expect differences between coders and analysts to affect a smaller portion of the sample than differences between interviewers. Furthermore, it will be assumed that response variance for establishment surveys conducted by mail is due primarily to (1) that the same respondent does not answer for the establishment at each survey occasion (and the associated variability between these respondents) and (2) transcription error and other errors by respondents which may or may not appear over repeated trials. Response bias is primarily due to the recordkeeping system and the average respondent's (over all possible respondents at the establishment) interpretation of the questions on the survey instrument as it relates to the recordkeeping system of the establish-

ment. Hence, the MSE under the authors' assumptions, is:

$$\text{MSE}(y) = 1/n \ (S_\mu^2 + \sigma_R^2) + B^2.$$

Furthermore, it will be assumed that in establishment surveys with CATI collection, the interviewers will not contribute substantially to correlated response variance because the respondent is only reading answers from a previously completed survey questionnaire. Also, the edit reconciliation questions are available to CATI interviewers.

Given the above model, traditional edit and screening approaches are designed primarily to identify simple response variance, especially in panel surveys where data are compared over different reporting time periods. For example, CATI on-line editing in the CES instrument includes internal consistency, range, and longitudinal checks. Internal consistency edits determine whether impossible combinations occur, such as the number of women workers exceeding total employment. Range checks determine if ratios of data items or calculated data, such as average hourly earnings, are reasonable for reporters with similar characteristics. Longitudinal edits compare the reported data to the previous month's data. These types of edits are common in panel surveys collected by mail, as well as by CATI.

All three types of edits are designed to detect response variance. Only the range checks are designed to detect response bias. However, range checks are usually too broad to detect many definitional inconsistencies. For example, a range check for average hourly earnings for production workers in a particular industry would most likely not detect the exclusion of both overtime hours and pay. The range checks are intended to detect data which are unreasonable relative to similar firms (by industry and size), and therefore must be broad enough to accommodate normal differences among firms.

The independently conducted Response Analysis Survey (RAS) is designed to identify and correct for bias that may be present in ongoing monthly establishment surveys. The RAS educates the respondent on survey definitions, and relates the definitions to the employer's recordkeeping system. Hence, bias will theoretically be reduced in periods subsequent to the Response Analysis Survey. The effect of the Response Analysis Survey on bias was measured in two ways. First, separate estimates were made of the characteristics of interest for the test group (those participating in an RAS) and the entire CES sample before and after conducting the RAS. Second, indirect measures of bias were obtained for each data element by tabulating the types and number of adjustments necessary and adjustments agreed to for the RAS sample.

2.4. Description of the record check approach

Each CES form consists of a space to record data and a set of definitions, consisting of a series of inclusions and exclusions for each data item, to aid the respondent in data preparation. However, there is no guarantee that respondents use these definitions.

The CES RAS instrument is designed to review the survey definitions against the firm's recordkeeping system to identify differences in definitions, and request that the discrepancies be corrected in the future. The instrument covers the primary sources of misreporting. The questions are neutrally worded seeking "yes-no" answers. A sample question is: "When you report total employment, do you include salaried officials?" If the respondent does not include such employees in the data reported, the respondent is asked if future data could include this

Fig. 1.   Response Analysis Survey percent of establishment adjustments by data item
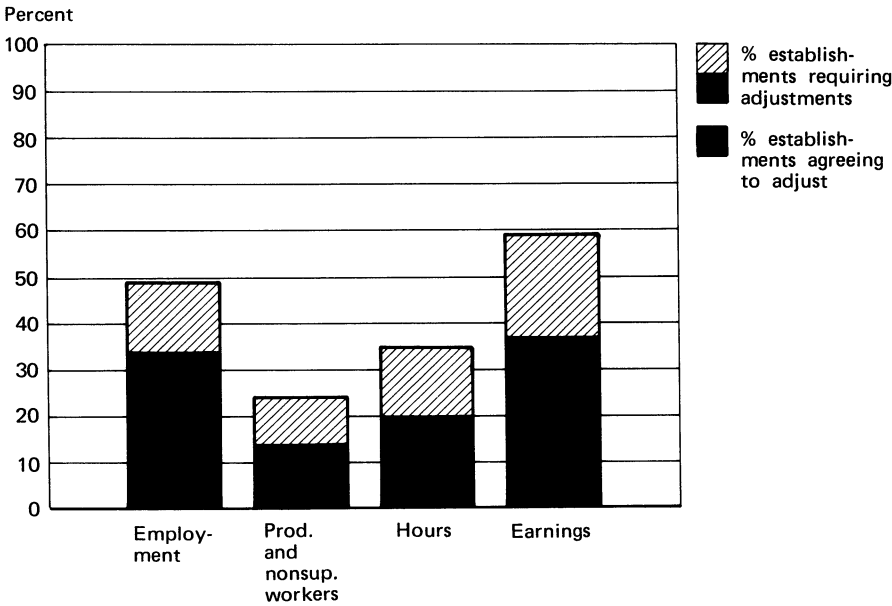
group. If a group is incorrectly included, the respondent is asked to exclude the group in the future.

As a means of perpetuating correct reporting, a form is sent to the respondent listing changes which the respondent agreed to make. A package is provided for both the collection log and the list of corrections for easy reference.

The Response Analysis Survey provides information about the survey concepts. If few respondents are able or are willing to report correctly, the survey practitioner may be forced to reevaluate the appropriateness of the concept or the approach being used. The reevaluation could result in changes in questionnaire data items, wording of questions, or definitions of data items in order to accommodate recordkeeping systems used by respondents.

Pilot tests were conducted in two states through personal visits which showed that respondents were willing to conduct such interviews lasting 10 to 15 minutes and that they were willing to discuss details of their record-

keeping practices. Due to the size of the sample (300 000) and the high cost of conducting personal visits, research was conducted on the feasibility of CATI to control response errors. The results have shown CATI to be as accurate as the personal visit and economically feasible for large scale implementation as a permanent quality control feature.

2.5. Results of the record check approach

Survey practitioners often depend on the survey reporters to adhere to definitions supplied on the survey instrument. Evidence from the series of record check studies conducted in conjunction with the Current Employment Statistics Program suggests that survey practitioners should be advised of the potential for large deviations from survey definitions. The Current Employment Statistics Program has a short list of items, for which the definitions appear straightforward. However, over half of the establishments studied were not adhering entirely to the survey definitions, primarily due to the recordkeeping system at the estab-

Table 1.   Response Analysis Survey impact of RAS on data items

	Average percent change Sep.85 to Sep. 86	
	RAS units	Control group
Employment	1.3 (3.9)	4.8
Women	6.0 (2.4)	6.9
Prod. Workers	5.3 (2.2)	5.8
PW. Hours	8.8 (4.2)	6.7
PW. Earnings	10.7 (3.2)*	1.6

( ) Standard error
* Significantly different from control group at 2 SE's

lishment. The number of deviations from survey definitions are referred to as "adjustments needed" in Fig. 1. The extent of the number of adjustments varied little over type of establishment (size of establishment, as well as industry).

An indirect indicator of the impact of the RAS was obtained by comparing the reported data for the same month a year later, for both the RAS sample and the entire CES sample for those states in which the RAS was conducted. The average changes (in percent) in reported data are given in Table 1. The RAS sample showed a significantly larger change in production worker earnings than did the entire CES sample. There were no significant differences for the other characteristics. Examination of average changes for the RAS sample based on the expected direction of the change indicated that units changed their reporting habits. While the sample sizes were too small to produce significant results, the results were consistent with the supposition that the adjustments were large relative to economic changes.

2.6. Areas for future study

The results indicate the ability of the RAS to identify a variety of response errors. Future studies will cover the following four areas.

1. Further assessment of the affect of response errors before and after the RAS will be conducted. Direct error measurement on components of error will help strengthen the RAS and provide an error profile for the survey.

2. There are other potential uses of this approach for addressing response errors in the CES program. For example, selected questions covering the predominant sources of error could be included periodically in the monthly CATI collection interview. These would ask whether the reported data are properly adjusted. Similarly, mail respondents may be sent a letter with the survey instrument which requests the respondents to ensure, for example, the inclusion of officers and executives in the "all employees" count.

Another potential use of the record check concept addresses evolving financial issues. An example is in the concepts of payroll. Many industries are providing employees with lump-sum payments as a part of the overall compensation package. This growing practice has implications for wage escalation contracts, one of the primary uses of the CES payroll data. The BLS definition excludes this sort of payment because inclusion of such payments would distort the published estimates. The record check approach would be useful in industries where lump-sum payments are prevalent. The CATI data collection could include a question checking that these payments are excluded, although the effect of such probing on ongoing response rates would be monitored. Similarly, questions could be added periodically to the CATI questionnaire to assess or evaluate other specific errors.

3. Procedures will be continually evaluated and refined. The method for perpetuating improvements will be reviewed and improved. The list of changes which the respondent agreed to make is sent to the respondent. Future tests may include a separate list of changes which were identified, but for which there was no consent to change. The respondent will be asked to make the additional adjustments when possible. For example, if the firm's payroll recordkeeping system changes, the respondent may be able to make some adjustments not now possible. In this way, the interviews may yield both immediate improvements and establish the prerequisites for future improvements.

4. Analysis of the effect of these methods on the estimates will be conducted, including

follow-up studies to assess whether the improved reporting is long-lasting, and thus, how often reinterviews should be scheduled.

3. CATI and Self-response Using Touchtone Recognition Technology

3.1. Background

The previous section dealt with the use of computer-assisted collection methods for the control of response error. In this section we will look at the application of computer-assisted collection methods for the control of sources of nonresponse error, including both CATI and touchtone self-response (or TDE for touchtone data entry).

Nonresponse can arise from a variety of sources in establishment surveys. For example, employers are often reluctant to agree to participate in the survey. Other error sources are: sample attrition over subsequent reporting periods; late reporting due to respondent neglect; unavailability of records or ineffectiveness of the collection mode.

There have been three primary modes of data collection for establishment surveys: personal visit, mail, and telephone. Personal visits can be important for sample initiation to obtain high response rates and high quality data, particularly for complex data elements on questionnaires. However, its practical utility in ongoing surveys is limited to smaller surveys with lengthier collection cycles due to its cost and collection time. Mail collection has been the dominant mode of collection for establishment surveys due to cost considerations and the need for employers to assemble their firms' records prior to responding. However, while being economical, mail collection has built-in time delays (mailout and mail-back) and quality limitations which make it a potentially unreliable mode of collection for surveys having tight, fixed publications deadlines. In the past, telephone collection has offered advantages for timeliness of collection, but has also been subject to potential for quality loss. The role of telephone collection in establishment surveys was that of a secondary collection vehicle primarily used for nonresponse prompting in mail surveys and edit reconciliation callbacks.

Over the past few years, major enhancements in the area of computer-assisted telephone interviewing have appeared, offering broad flexibilities in survey management, data quality, timeliness, and quality control. CATI addresses many of the past quality concerns with the use of telephone interviewing while allowing data collection to become more centralized, controlled, and cost effective. In the following sections, we will discuss the results from CATI feasibility tests on the control of nonresponse error sources in the monthly CES survey and the extension of these computer-assisted techniques to a more cost efficient employer self-reporting mode using a BLS developed touchtone data entry system.

3.2. Shifting from mail to CATI

The current collection mode for the CES survey is via mail. Each month, the respondent receives the collection schedule and fills in the relevant data. The form is mailed to one of 51 State Employment Security Agencies where the mail is processed through key entry, verification, editing and reconciliation. The preliminary estimates are based on the data collected during a two-week period. Responses from approximately 45 % of the sample are collected during these two weeks; these data are then sent to Washington, D.C. for processing. The same procedures are used on the remaining sample, resulting in two subsequent publication releases.

The CES Program relies on the states to conduct all survey operations under specifications from the national office and technical guidance from eight regional offices. Each

state is a separate legal entity with varying bureaucracies, procurement laws, personnel systems, etc.

The strength of collection by mail is that it is a relatively inexpensive and simple collection process. However, the weakness is the lack of control over the timeliness of the mail collection process and the very short processing "window" available for edit reconciliation and nonresponse follow-up. Respondents must wait for the payroll records for the reference period to become available before completing the form. The combination of widely different payroll periods – weekly, bi-weekly, semi-monthly, and monthly – and other respondent delays present problems in timeliness and balance of the preliminary sample under the mail process.

To address these problems, small scale feasibility tests of telephone collection were initiated in 1984. These were followed by tests using CATI in 1985. The results of these tests using representative samples in two states were very promising. To complete the interview, the average respondent required only 1.5 calls lasting a total of about 3 minutes. Also, 90 % of the sample units were collected for the preliminary estimates versus 45 % for mail.

These results led to development of production tests, the goals of which are to conduct CATI collection on a large enough scale to devise a cost-effective implementation strategy. Information is being collected on: CATI performance (to verify earlier results), costs, and the effect the change to telephone may have on the state organization. Nine states and over 3 800 respondents are part of the CATI testing. The CATI samples in six states are designed to be representative of the CES samples. In the remaining three states, CATI samples are drawn from the least timely reporters to test the limits of CATI's ability to improve timeliness.

The first phase of the production tests has been completed. Preliminary results verify those from the earlier tests: 1.5 calls per respondent totaling about 3 minutes; between 85 and 90 % of units are collected for the preliminary estimates. Further studies point to other benefits of CATI. Sample attrition is one-half that of mail and item nonresponse appears to be reduced. Thus, primary indicators point to CATI as a more effective collection mode than mail. In addition, a great deal of information is obtained through the CATI interview which is not obtainable with mail data collection. For example, reason for nonresponse (out-of-business versus refusal) is available on a timely basis for the CATI sample, but not for mail. Full analysis of these results will not only provide a clearer picture of CATI performance, but will also provide more complete information about the limitations of mail surveys. The second phase of the production tests will include a doubling of the CATI sample comprised solely of late reporters. However, while CATI collection demonstrated many strengths in resolving current problems in data collection, preliminary data indicated that the cost of full CATI could not be absorbed within the current survey budget.

In order to maintain the performance gains experienced under CATI, research was initiated into the area of touchtone data entry to develop an alternative telephone collection method with a lower unit cost.

3.3. Description of the integrated TDE-CATI system

The use of touchtone recognition technology for surveys is initially met with some skepticism by survey practitioners. The idea of respondents reporting their data to a computer simulated voice conveys images of large refusal rates; however, current research results suggest that the use of touchtone recognition technology and computer simulated voice may have uses in a range of survey applications. In particular, for panel surveys where

the respondent is reporting the same data items monthly or quarterly, self-reporting by touchtone phone can be a convenience to the respondent. The respondents are given the option of calling in their data any time, day or night. The units selected for touchtone reporting in the CES survey have all been regular CATI reporters, and are therefore in the habit of preparing their data to report to a telephone interviewer each month.

Features of the touchtone self-response system include:
– the detection of legitimate reporters based on a match to a file of reporter numbers;
– the detection of reporters which have called in their data more than once (e.g., the respondent detects an error after the initial call);
– varying the set of questions depending on the industry of the reporter;
– reading back all responses for respondent verification using computer simulated voice; (e.g., the respondent enters a "3" using the touchtone phone, the computer responds with "you are entering data for March"; the respondent can then confirm this with a "1", or enter "0" to try again);
– limited interview oversight (e.g., the respondent has three opportunities to report any data item correctly, and is required to confirm all entries); and
– storage of the date, start time, and end time of each call.

The system being tested requires a dedicated microcomputer, and boards which allow multiple phone lines.

The primary reason to consider a touchtone self-response survey application is to reduce the cost of collecting data by CATI, while maintaining the quality and timeliness of CATI collection. Hence, in designing the system it was necessary to integrate the self-response system with the CATI system. First, all units which fail to self-report their data within a given period would receive a CATI call. Second, all reported data are processed through the CATI instrument and edit checks. Units failing edits would receive a CATI call to resolve the edit failures. There has been some consideration of editing the data while the units are self-reporting their data. It was decided that resolving edit problems using the computer simulated voice could jeopardize response. However, the decision to incorporate and test edit-resolution questions in the touchtone instrument is still under consideration.

The integrated touchtone self-response and CATI system has the ability to identify units which:
– have reported their data on a daily basis;
– are delinquent in reporting and will, therefore, require a CATI call; and
– have reported their data and have failed one or more CATI edits.

Units failing one or more CATI edits will also receive a CATI call unless the interviewer has evidence of the reason for failing edits, such as normal seasonal fluctuation for this particular reporter. All other units have presumably reported their data, and have passed the CATI edits. The data for these units will be transferred directly to the CES data base for processing.

The initial test of this system was primarily designed to determine the:
– conversion rate from CATI to TDE;
– TDE attrition rate versus CATI and mail rates;
– number of units requiring prompting calls to assure timeliness;
– percent of units requiring edit reconciliation calls;
– cost-effectiveness of an integrated TDE-CATI system; and
– needed changes in procedures, hardware, and software.

The number of units for which data could be directly transferred to the CES data base provides an indication of the potential for cost savings of the integrated TDE-CATI system over a CATI only system.

3.4. Results

In 1987, TDE tests were conducted in two states to determine the operational feasibility of this technology for the Current Employment Statistics Program. The feasibility tests were designed to provide limited performance measures, as well as to allow continued development of the system and procedures for state personnel. The feasibility tests were small in scope (about 200 monthly reporters split between the two test states). The performance measures obtained in this test were encouraging, and a larger scale test in four states was undertaken in 1988.

In 1988, the TDE tests were expanded to over 600 monthly reporters. A number of enhancements were made to the system, including mailing a monthly postcard reminder to each reporter. CATI prompt calls were placed to nonrespondents about three days before the cutoff date for preliminary estimates. With proper control, it appears that between 60 % and 70 % TDE reporters will call in before receiving a CATI prompt. In addition, response rates for preliminary estimates are consistently over 80 %. The CATI prompt calls are short, and remind the reporters to call their data into the TDE system. Only about 10 % of reporters which were prompted needed a second call to collect the data. Hence, the burden on the data collection staff is greatly reduced, as compared to CATI.

The time of day reporters call in their data appears evenly distributed across "core" working hours (9 a.m. – noon, 1–4 p.m.). The number of phone lines needed for a given sample size is being analyzed based on the date and time data.

3.5. Future activities

A major obstacle to large scale TDE use is the requirement that the establishment has access to a touchtone telephone. While touchtone availability is increasing, an exten-

sion of self-response is being pursued using voice recognition technology. Thus, an inexpensive method of reporting would be available for any respondent with a telephone. Voice recognition technology is improving rapidly, although its application to data collection will require rigorous testing.

This TDE-CATI implementation partitions the sample in two groups. The first group is large reporters of 500 or more employees. These reporters represent about 5 % of units and over 60 % of unweighted sample employment, and are being considered for permanent CATI collection. Thus, CATI is being focused on an important component of the sample and one which is traditionally difficult to collect in a timely manner.

The second group consists of all establishments with less than 500 employees. This group represents 95 % of units and about 40 % of the unweighted sample employment. Those units which have difficulties in terms of timeliness of responses will be addressed first. Panels of poor reporters will be converted to CATI for several months, in an attempt to improve their reporting habits. After a reporter has demonstrated consistently improved reporting behavior, TDE will be offered. Then another panel of poor reporters will be rotated through CATI towards TDE. Over a period of a few years, a growing component of sample units will be rotated into TDE. The anticipated cost savings will provide a basis for the more intensive CATI follow-up procedures. In addition, through the rotational procedure an increasing proportion of the sample will be included in preliminary estimates of employment, hours, and earnings.

4. Summary

The two survey methods discussed in this paper have been shown to be effective tools for the Current Employment Statistics Program. Several thousand Response Analysis

Surveys are currently underway, and the touchtone project is expanding during 1989.

We believe that other surveys are potential benefactors of these methodologies. The response profile presented for establishment versus household surveys indicates that establishment surveys also have a large potential for response error, although not necessarily for the same reasons as household surveys. In the past, little attention has been paid to research in response error problems for establishment surveys. Our current record check survey research has confirmed the existence of problems and the need to exercise ongoing control over response error sources in establishment survey data. A record check approach for pretesting surveys is not new, but is being used more frequently. We would encourage conduct of these surveys and publication of results to provide survey practitioners with additional design ideas.

The current Employment Statistics Program is testing touchtone self-response, CATI, and mail collection as an effective and economical mixed mode data collection utilizing the strengths of each mode for the relevant portions of the sample. We also encourage the development and testing of touchtone self-response in other surveys as a method to improve the timeliness of response and reduce costs.

Touchtone self-response applications for surveys are wide and varied. Many household and establishment surveys could consider touchtone self-response as one of a mixed mode data collection. Certainly for many other panel surveys it may be viewed as a convenience to respondents. It may also be appropriate for quick response surveys, where a questionnaire is mailed to respondents, who are then given an option of self-response by touchtone phone, resulting in time and keypunch savings. Finally, touchtone recognition can be an effective management information tool in a decentralized organization.

Interviewers or field representatives can report completed cases daily by touchtone, thus providing progress reports to management.

BLS research will continue to make use of technological advancements in survey operations. With increasing pressure to do more with less resources, we need to take advantage of technological advancements such as touchtone recognition, voice recognition, and other advances to maintain or improve the quality of our existing surveys.

6. References

6.1. *References cited in the text*

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1961): Measurement Errors in Censuses and Surveys. Bulletin of International Statistical Institute, 38, pp. 359–374.

United States Office of Management and Budget Statistical Working Paper 15 (1988): Quality in Establishment Surveys.

Werking, G.S., Clayton, R.L., Rosen, R.J., and Winter, D.S. (1988): Conversion From Mail to CATI in the Current Employment Statistics Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods.

6.2. *References not cited in the text*

Groves, R.M., and Nicholls, W.L., II (1986): The Status of Computer Assisted Telephone Interviewing: Part II – Data Quality Issues. Journal of Official Statistics, 2(2), pp. 117–134.

Werking, G.S., Tupek, A.R., and Clayton, R.L. (1988): CATI and Touchtone Self-Response Applications for Establishment Surveys. Proceedings of the U.S. Bureau of the Census Annual Research Conference.

Werking, G.S., Tupek, A.R., Ponikowski, C.H., and Rosen, R.J. (1986): A CATI Feasibility Study for a Monthly Establishment Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods.

Werking, G.S. and Tupek, A.R. (1987): Modernizing the Current Employment Statistics Program. American Statistical Association, Proceedings of the Business and Economics Statistics Section.