

Data Collection Issues Related to Implementing the Redesigned Current Employment Statistics Survey
Richard J. Rosen, Christopher D. Manning, Louis J. Harrell, Jr., and Douglas A. Skuta

Christopher D. Manning, BLS, 2 Massachusetts Avenue, N.E., Ste. 4860, Washington, DC 20212
Manning_C@BLS.GOV

Disclaimer: Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics

Key Words: Probability-based sample, Quota sample, Establishment Surveys

Introduction: In June of 1995, the Bureau of Labor Statistics (BLS) announced plans for a comprehensive redesign of the Current Employment Statistics (CES) Survey sample. Based on several years of research and the recommendations of a number of expert panels, BLS launched sample solicitation in mid-1997 for a new probability design to replace the existing CES quota sample. The new sample design is a simple random sample of Unemployment Insurance (UI) accounts.

Initiating the new sample requires soliciting ongoing reporting from over 240,000 firms. This large-scale undertaking required development of new enrollment protocols and procedures. Initiation of the new sample is accomplished primarily via Computer Assisted Telephone Interview (CATI). Ongoing data collection utilizes such automated methods as Touchtone Data Entry (TDE), FAX, Electronic Data Interchange (EDI), and World Wide Web.

In this paper, we discuss the issues involved in implementing data collection in the redesign sample. We focus on developing the necessary infrastructure to perform enrollment and data collection, development of respondent materials, and management of survey activities across three Data Collection Centers, other sites, and multiple collection modes. Results to date and implications for other establishment surveys will also be discussed.

Background on CES Program: The CES Program is a monthly survey of about 380,000 business establishments. The CES Program provides data on employment, hours, and earnings by industry and geography. The CES operates in a Federal-State cooperative system where each state collects, enters, edits and transmits data for the national estimates. CES data, published after only two and a half weeks of collection, are widely viewed as a major economic indicator and serve as a key measure of the health of the economy.

The CES program offers several important features to its data users: timely release of data, an abundance of industry and geographic detail, and an annual benchmark to full population counts from state Unemployment Insurance (UI) tax records, which helps to maintain overall survey accuracy. However, the current CES sample design suffers from two main limitations: the lack of a probability-based sample design, and the absence of a method to directly measure new business births. Both limitations have been addressed in the redesign sample.

The current CES sample is a quota sample that was established in the 1920s, before probability sampling was internationally recognized as the standard for sample surveys. Over the years, however, both internal and formal external reviews of the CES program (Gordon Commission in 1960, Levitan Commission in 1980, and an ASA expert panel in 1993) concluded that probability-based sampling would benefit the program. The current sample also employs a limited and elementary modeling procedure to account for the presumed missing employment from birth units.

The new sample design (referred to as the redesign sample) is a simple random sample of Unemployment Insurance accounts. Since quota samples are known to be at risk for potentially significant biases, introducing a probability-based sample for CES will more effectively ensure a proper representation of the universe through randomized selection techniques and the regular rotation of sample members. The design also addresses the problem of new business births by systematically adding new births to the sample. This improved birth/death measurement, coupled with a probability-based sample design, should yield more accurate and consistent employment, hours, and earnings estimates across the national, state, and area levels.

Data Collection Infrastructure: Mixed mode data collection is essential to the success of the CES program. The survey is voluntary, covering geographically dispersed firms varying in size, and we must be able to accept data in any format that the respondent wishes to use. The redesign sample is

collected at five different sites using a variety of collection methods. Sample enrollment and initial CATI collection are conducted in Data Collection Centers (DCCs) located in Atlanta, Dallas, and Kansas City. A site in Chicago uses Electronic Data Interchange (EDI) to collect data from very large, select firms. A centralized Touchtone Data Entry (TDE) facility in Washington, DC, is used to collect data from respondents that have been transitioned from CATI reporting. A small number of cases are collected using the World Wide Web.

The locations of the Data Collection Centers were selected because of the availability of both interviewers and experienced survey managers. A Manager, Assistant Manager, four supervisors, and 40 interviewers staff each DCC. Interviewers usually have backgrounds in telemarketing or customer service and receive training on using the CATI software, on the sample enrollment protocol, and on the treatment of reluctant or refusing respondents. The EDI center is staffed with a Manager and six staff members. The One Point TDE facility has a team leader, three staff members, a Help Desk supervisor, and six Help Desk specialists.

The DCCs use CATI software developed by BLS to enroll respondents and to collect data. The first version of the software was fielded in 1996, and has been continually refined to meet interviewer and program needs. We are now using the third version of the software, released in October, 1998. It is a modified client-server system implemented using Visual FoxPro and NT servers. The software's design is robust; interviewers are able to collect data without worrying about the effect of server downtime. Supervisors allocate cases from the server to individual interviewer computers, and at the end of the day, completed cases are moved back to the server for uploading to the estimation system.

The EDI software was developed in 1994. It is designed for collecting data from very large companies. Participating companies usually have hundreds of locations and report information for all locations in a standard format. Data are automatically generated by the firm and electronically transmitted to BLS. The CES program is using EDI to collect data for the current quota sample from 15 large firms representing 2.4 million employees and 21,700 establishments. Because of their large size, these large firms were also selected for the redesign sample. In addition, BLS has enrolled four additional large firms needed for the redesign sample.

The backbone of CES data collection is Touchtone Data Entry. TDE is used for both the old quota sample and the new redesign sample. Approximately 250,000 reports are collected through this mode, most in the current quota sample. Respondents using TDE call a toll-free number, then enter their data using the number pad of their telephone. The software is written in C++. The TDE hardware includes computer telephony cards manufactured by Dialogic and Windows NT servers.

The TDE system offers significant advantages over most traditional collection methods. It combines low cost of operation with the ability to use broadcast FAX to perform "Advance Notice" and nonresponse prompting (NRP) functions. TDE has been able to sustain ongoing response rates in the 70-75% range.

Web based data collection is also used for a small number of units. We have been testing web based data collection since 1995 and are currently collecting data for the current quota sample from 100 respondents on the Web. Redesign units that decline to report through TDE are offered the option of Web reporting. A few respondents have specifically requested Web reporting.

The Web collection system has been highly successful. It is the least expensive collection mode and is obtaining response rates similar to those obtained through TDE. The constraint to its use is the relative lack of availability of Web and e-mail access on the desktops of our respondents. However, this situation is rapidly changing (Harrell, 1996).

Development of Enrollment Protocol: In 1996, BLS and the University of Michigan Survey Research Center conducted research on a sample enrollment protocol. The goal of the research was to develop a protocol that optimized response rates and could be easily implemented (Groves, 1997).

The research consisted of six experiments. The first experiment looked at the amount of contact information that could be collected for address refinement purposes. Another project addressed the appropriate level of entry into the firm for obtaining compliance with our request. A third experiment assessed the impact on response rates of varying the requested length of participation. Other tests looked at the effect of emphasizing State mandatory reporting laws, redesigned respondent materials, and varying the length of time before transferring a case to TDE collection.

The results of the tests were used to refine the protocol that we adopted for our live production tests. In most instances, the results of these tests did not demonstrate significantly superior treatment effects. However, these tests provided an enhanced set of respondent materials for use in our production tests. Another major initiative was the development of a special training module to address respondent reluctance and assist in "refusal aversion". This workshop has been very effective in convincing reluctant respondents to participate.

Implementing the Redesign sample: Each month, Data Collection Center interviewers receive a panel of 50 new cases to process. These cases progress through three distinct but interrelated phases: address refinement, enrollment, and data collection.

Address Refinement: The objective of the Address Refinement phase is to locate the firm's contact person as well as address and phone number. The DCCs usually conduct Address Refinement for all new cases during the first week of the collection period.

The Unemployment Insurance (UI) frame provides basic information about each sample unit. However, the UI frame is 15 months old when the sample is fielded. Therefore, the UI information must be updated and supplemented in order to enroll the unit. In most cases there is some vital information missing that is needed to enroll the unit. For example, only 67% of the UI frame has a telephone number, and less than 5% has a contact name. Therefore, Address Refinement usually consists of (1) verifying or obtaining the establishment's phone number, and (2) calling the establishment to verify the address information and to obtain the name of the person most likely to become the CES contact person.

Interviewers must be quite resourceful in obtaining or verifying the phone number. Tools used to update and complete the address information include directory assistance, commercial business directories, internet search engines, city and local government organizations, and Secretaries of State. When necessary, interviewers will even use less conventional methods such as calling an adjacent business. Once a valid phone number has been obtained, interviewers will usually call the establishment to verify the address information and to verify or obtain the name of the person most likely to become the CES contact person. If no contact name is provided from the UI frame, interviewers normally ask for the head of payroll, the office manager, or the business owner.

During the refinement call the interviewer emphasizes their affiliation with the Department of Labor and mentions that some materials will be mailed within the next few days. Interviewers do not request participation in the survey at this time. Once the refinement call has been completed, an information packet is mailed to the prospective respondent and an enrollment call is scheduled for 5 to 7 business days after the mailout of the package.

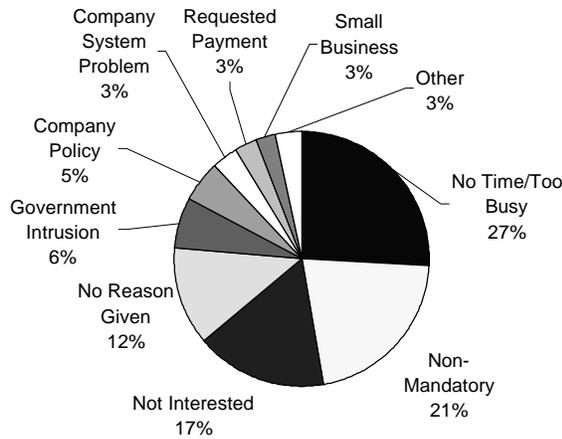
Enrollment: The objective of the Enrollment phase is to obtain the respondent's participation in the survey. Successful enrollment of new respondents is critical to the success of the redesign initiative. To that end, the CES must be presented in such a light to elicit participation in the voluntary survey. The information packet is the first step in this process.

The specially developed information packet is mailed to the prospective respondent at the end of the address refinement phase. The packet includes a customized letter requesting participation which is signed by the DCC manager, an informative brochure presenting an overview of the program, a fact sheet explaining how the requested data items are used, and an industry-specific CES data collection form. Approximately 5 to 7 days after the packet is mailed, an interviewer will call the potential respondent to begin the enrollment process. During the enrollment call, interviewers use strong positive enrollment techniques in an effort to obtain the respondent's participation. Interviewers attempt to establish a dialog and build rapport with the contact person while stressing the importance of CES and the ease of reporting. They inquire about the receipt of the respondent packet, provide information about the survey, and answer questions and concerns that the respondent may have. Once they gain participation, the interviewer schedules a callback to collect data and the case progresses to the data collection mode.

Interviewers are trained on refusal aversion techniques and apply them when talking to a reluctant respondent. "Refusal aversion" refers to methods used when encountering initial refusals in an effort to turn them into respondents. A training class was developed to build interviewer skills in overcoming refusals. The class, which covers three days, introduces interviewers to the major reasons for refusing to participate and provides possible responses to the objections. Exercises allow the interviewers to improve their expertise in rapidly overcoming objections. As Figure 1 shows, the most significant reasons for refusing to participate are the lack of time (27%), the non-mandatory nature of the

survey (21%), and a general disinterest in participating (17%).

Figure 1. Most frequent reasons for refusing



If interviewers encounter initial reluctance to participate, they are trained to actively listen to the respondent's issues, show empathy towards their concerns, identify the real barrier to participation, and offer a counter response that appropriately addresses the reason for reluctance. If they still encounter reluctance, they may pursue avenues such as reducing the number of data items requested or accepting an aggregated report such as countywide or statewide (if the UI account contains multiple establishments). If the reluctance continues at this point, the interviewer breaks off the interview in a friendly manner. A more experienced interviewer will then call the respondent back at a later date and attempt to convince them to participate. This is referred to as the 'refusal conversion' call.

Table 1. Success rate of refusal conversion

Refusal Reason	Percent Converted
No reason given	75%
Other	69%
Small business	60%
No time / too busy	41%
Requested payment to report	40%
Company systems problem	38%
Not interested	29%
Company policy	24%
Government intrusion	21%
Non-mandatory	15%
Total	35%

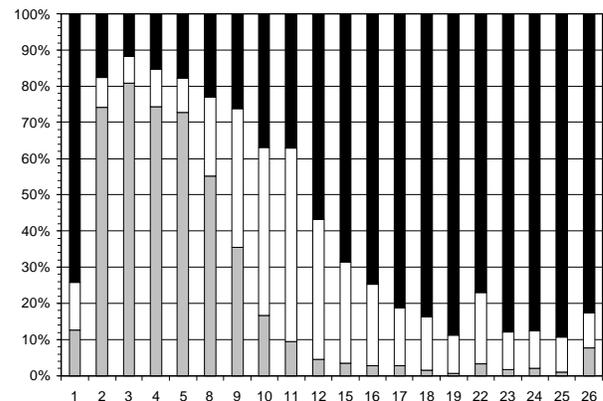
Table 1 shows how successful we have been in our refusal conversion efforts. As might be expected, respondents who were 'hard' refusals--those who

cited "non-mandatory", "government intrusion", and "company policy" as reasons for refusing--were the most difficult to convert. It may be worthwhile to first focus conversion efforts on 'soft' refusals. We seem to have more success converting those respondents who cited reasons such as "no time", "too busy", "not interested", "small business", and "requested payment for reporting", or who gave no reason at all. Overall, we are able to 'turn around' 35% of all initial refusals.

Data Collection: The Data Collection phase has a dual objective: to actively collect data while preparing the respondent to report using a self-reporting automated collection method. Initially, interviewers call respondents at a mutually agreed upon time for 6 months and collect their data during a CATI interview. Data are automatically edited upon being entered by the interviewer. Data that fail the edit screens are reconciled during the CATI call. The bulk of the data collection calls are made during the last two weeks of the collection period.

Figure 2 displays a snapshot of the DCC workflow during a typical month where each bar represents one working day. Within each day, the color differences represents one of the three major activities performed by interviewers (address refinement, enrollment, and collection). The black bars denote the monthly data collection workflow. This activity is a constant throughout the month; however, collection virtually drives out address refinement (gray bars) and enrollment (white bars) activities as the collection deadline approaches.

Figure 2. Workflow in the DCCs



It is during the data collection period that interviewers educate the respondent about what data items are needed and attempt to help the respondent

establish a good reporting pattern. This is important because the respondent will soon be asked to report using a self-reporting automated collection method such as TDE, FAX, or World Wide Web (WWW).

TDE is the first option for respondents who have completed 6 months of CATI collection. During the 5th month of CATI collection, interviewers will provide a brief overview of TDE to the respondent and inquire about the ability of the respondent to report using a touchtone telephone. Those who are eligible and receptive to Touchtone reporting will receive a TDE information packet prior to the 6th month of CATI collection. The TDE information packet includes a customized letter requesting that the respondent report via TDE, a fact sheet explaining how the requested data items are used, an instruction sheet explaining how to use the TDE system, and an industry-specific CES data collection form. During the 6th month CATI collection call, the interviewer will verbally explain the procedures for TDE reporting and formally request that the respondent begin using the TDE system to report the following month. Respondents who do not oppose are converted to Touchtone reporting. Respondents who are not able to report via Touchtone are given other options for automated reporting, such as WWW or FAX. Some remain on permanent CATI collection.

Following the 6th month CATI collection call but prior to the 1st month on TDE, interviewers will make a "Ready to Report" phone call to the respondent. The purpose of this call is two-fold: (1) to ensure receipt of the TDE packet, and (2) to remind them to report via TDE. During this call, interviewers attempt to answer any remaining questions or concerns that the respondent may have about reporting using TDE.

One important responsibility of the interviewer during both the enrollment and data collection phase is to attempt to identify establishments that have recently opened or closed. The estimator for the redesign sample, a weighted link relative type, requires us to identify new locations and locations that are closed. Procedures and training were developed to implement collection of the new locations. Interviewers are instructed to query respondents during every interview on changes in reported establishments, and the CATI software allows the interviewers to enter new locations and collect data for them, as well as allowing them to code closed locations as 'out of business'. The TDE software was upgraded to allow respondents to notify BLS of an opening, closure, merger, or acquisition

affecting their report. If a respondent indicates that a change has occurred, an interviewer will call them to verify the nature of the change and to collect information on initial employment and location of the new establishment.

Address Refinement, Enrollment, and Data Collection activities constitute the bulk of the work done by the DCCs. However, interviewers are responsible for working on a myriad of other tasks. They also make "Advance Notice" phone calls to high risk reporters reminding them to report; "Non-response prompt" phone calls to respondents that have not reported and are past their expected report date; and "Long-term" non-response phone calls to ongoing sample units that have not reported for 3+ months.

Managing Data Collection: Management of data collection requires implementing several related processes. The sample allocation process and receipt of management information files are essential to CES data production and evaluation of CES operations. These processes have required the development and maintenance of several databases. Survey managers also rely on use of e-mail, conference calls, and bimonthly meetings for managing operations.

Fielding a new sample while collecting data from the old sample is challenging. The sampling unit in the current quota sample is the establishment. The sampling unit in the new probability design is the Unemployment Insurance (UI) account, which could include multiple establishments. Therefore, about 15% of the redesign sample overlaps the current sample.

Units that are in scope for the new design are: (1) the total overlap cases where we already collect data for all of the establishments in the UI account; (2) the partial overlap cases where we collect data for some of the establishments in the UI account and need the remaining part, and; (3) those firms that do not match at all to the current sample. The partial matches and no match cases are grouped by industry and allocated in panels to the Data Collection Centers.

We rely on several databases to correctly identify cases scheduled for allocation to a DCC. Each interviewer receives 50 new UI accounts for enrollment each month. The panel allocation process ensures that one interviewer is the contact with a particular firm. Thus, the Employer Identification Numbers (EIN) of previously allocated firms are matched to the monthly panel to ensure that the same interviewer contacts all parts of a firm. The file is

also reviewed to ensure that non-matching EINs that are part of the same firm are allocated to the appropriate interviewer. Finally, the monthly panel is sorted by State, so that each DCC only receives certain States.

A small number of these cases are excluded from the paneling process due to size or unique collection difficulties. The exclusions are large firms scheduled for enrollment through a personal visit, staff leasing firms, and cases where special reporting arrangements already exist. The excluded firms are characterized by their size and collection difficulty. Personal visit cases are very large firms in their industry that appear to offer some possibility of providing a centrally generated report to BLS. Regional Office staff visit these firms and attempt to enroll them into CES. The leasing firms report for a variety of industries and usually have hundreds of establishments. One leasing firm could easily provide a month's work for several interviewers.

Collecting management information is essential to monitoring the effectiveness of CES procedures. All data collection systems used in the CES survey automatically capture information about the data collection process. The CATI system captures a large amount of information about contact with the case. Information on the date, duration, type of contact, outcome, and other descriptive information about the case are stored in a text file that is extracted on a weekly basis. A polling agent, located in BLS headquarters, moves the information from the regional DCCs to BLS. The management data file has become a key resource for analytical studies of the results of data collection, and is used on a daily basis to answer questions and solve problems.

Results of Enrollment and Collection Efforts: We experienced a learning curve as the first industry, Wholesale Trade, was fielded. Software limitations and the lack of a refined enrollment protocol hindered the enrollment process. However, as systems and processes improved, so did results--particularly response rates.

Recent response rates have been quite high, reflective of strides that interviewers have made implementing the enrollment protocol. Response rates for current panels under control of the DCCs have averaged 68-70% across the 3 collection centers for the 5 month period April '99- August '99. The development of detailed tabulations based on the monthly system management information assisted in the tracking of DCC performance.

We also recognized the need to establish evaluation criteria and performance standards to measure individual interviewer workload and performance. These evaluation criteria have helped identify interviewers who are excelling in their efforts as well as interviewers who may need supplemental training. The performance standards and goals reflect program requirements and objectives. Interviewers are evaluated on three key performance measures: response rate, item response rate, and collection rate. These rates are defined in Table 2.

Table 2. Rates Definitions and Goals for Interviewer Evaluation

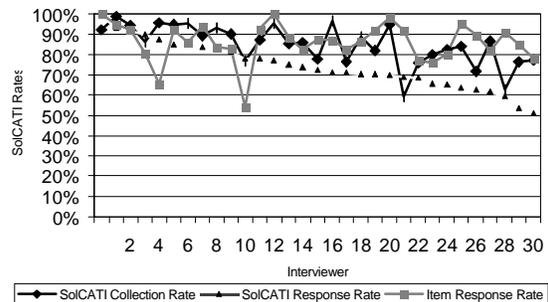
Rate	Definition	Goal
Response Rate	(# of units providing at least 1 month of usable data) / (total sample-OOB-OOS-duplicates)	75%
Data Item Response Rate	(# of response items)*(# of data reports) / (# of data items on form)*(# of data reports)	70%
Collection Rate	(# of units providing data) / (# of active units in collection mode-OOB-OOS-duplicates-refusals)	85%

OOB -- Out of business

OOS -- Out of Scope

Response and item response rates are comprehensive measures of response that are based on current "fully enrolled" panels. Collection rates measure the percent of active sample that provided data for the current month. Each month interviewers strive to meet a 75% response rate, 70% item response rate, and 85% collection rate. Monthly charts such as the one in Figure 3 assist us in identifying interviewers performing at outstanding levels as well as interviewers who would benefit from extra training.

Figure 3. Interviewer evaluation chart



Conclusions and Implications for Other Establishment Surveys: We have learned a great deal as we enrolled the Wholesale Trade and

Manufacturing industries. These lessons can apply to other establishment surveys as well.

Refining the enrollment protocol has allowed for better training of interviewers. Interviewers receive extensive training on a variety of issues, some of which were less of a focus before the implementation of the redesign sample began. For example, interviewers now receive extra training devoted to refusal aversion. The increased training has helped decrease the refusal rate as the implementation of the redesign sample has progressed from Wholesale Trade to Manufacturing.

Enhancement of the CATI software system has also improved results. As the implementation of the sample progressed, we learned more about the systems needs of the interviewers. These needs were built into later releases of the CATI software, and the improved systems have made interviewers more productive. For example, the current version of the CATI system has the capability to immediately FAX respondent materials to a prospective respondent. This reduces the number of calls required and impresses upon the potential respondent the importance of our requests.

CATI system management information has played a key role by allowing us to successfully examine workload and workflow issues. These workflow analyses were critical in defining DCC and interviewer productivity and performance measures. The analysis and review is a continuing process, both to measure performance and refine procedures.

Mixed mode data collection is vital to the success of the redesign implementation. Since the CES is a voluntary survey, we must be responsive to the changing needs of our respondents, and offering several different methods of collection is one way to meet these needs.

We will undoubtedly learn more as we field more industries. We have found that each industry has its own peculiarities that must be dealt with on an industry-by-industry basis. For example, some industries tend to have more establishments with employment of 1 that are difficult to contact, while others tend to report on a more aggregate level (less breakout by establishment), while still others have large staff leasing firms, which are a significant collection burden. It is these challenges that we will encounter as we continue to collect data for the redesign sample.

References:

Butani, Shail, G. Stamas, and M. Brick (1997). "Sample Redesign for the Current Employment Statistics Survey." *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 517-522.

Groves, Robert M. et al (1997). "Research Investigations In Gaining Participation From Sample Firms In The Current Employment Statistics Program" *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 289-294.

Werking, George S. (1997). "Overview of the CES Redesign" *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 512-515.

Getz, Patricia (1997). "Update on the Sample Redesign for the Payroll Survey", Internal Report, U.S. Bureau of Labor Statistics, Washington, D.C., (unpublished).

Harrell, Louis, R.L. Clayton, and G.S. Werking (1996). "TDE and Beyond: Data Collection on the World Wide Web", *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 768-773.

Rosen, Richard J., R.L. Clayton, and L.L. Wolf (1993). "Long-term Retention of Sample Members Under Automated Self-Response Data Collection", *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 748-752.

Werking, G.S., and R.L. Clayton (1991), "Enhancing Data Quality Through the Use of Mixed Mode Collection", *Survey Methodology*, June 1991, **17**, No. 1, pp. 3-14.