# Internet collection at the Bureau of Labor Statistics: an option to report data

*Most BLS programs offer employers multiple options for reporting data; the BLS approach to Internet data collection is to provide a single, manageable, secure architecture for all programs*

Stephen Cohen, Dee McCarthy, Richard Rosen, and William Wiatrowski

Many of the Bureau of Labor Statistics (BLS) statistical programs that collect data from employers, such as those reporting on employment, compensation, or occupational injuries and illnesses, are introducing Internet-based data collection. Among the reasons for offering Internet collection is the availability of reporting choices for employers, the potential to reduce BLS costs associated with mailing or data entry, and the potential to reduce employer burden, or the perception of burden. Multimode surveys that include the Internet allow respondents to select the most suitable mode for their circumstances, which may reduce burden. Internet-based data collection also augments the palette of the questionnaire designer by adding to it Internet technologies such as hyperlinks, color manipulation, dynamic graphics, and multimedia players that can provide instructions visually and/or orally.

This article reviews the BLS experience with collecting data over the Internet. BLS is in the process of incorporating Internet-based data collection into a number of its establishment surveys. Internet collection is typically offered to employers as one of several options for reporting their data. BLS benefits from Internet collection as well, with the potential to:

- Control certain costs
- Improve response rates
- Decrease burden, or the perception of burden
- Improve data quality
- For surveys with multiple deadlines—reduce revisions between preliminary and final estimates

The BLS approach to Internet data collection is to provide a single, manageable, secure architecture. Known as the Internet Data Collection Facility (IDCF), this facility has the following characteristics:

- A single entry point for all BLS surveys
- A common look and feel across surveys
- Support for multisurvey respondents
- Multiple levels of security
- A unique firewall, separate from the main firewall that protects internal BLS data
- Access controlled by BLS—issued accounts/passwords and digital certificates
- A single infrastructure, which helps to control monitoring and risk assessment activities

BLS is responsible for collecting data related to the workforce and is organized by program area, such as employment, prices, productivity, compensation, and occupational safety and health. The Internet collection initiative discussed here is available for data collected from employers.[1] Each BLS program that collects data from employers uses a variety of collection methodologies. For programs that are administered jointly with the States, initial collection is often by mail, with telephone followup of nonrespondents handled by State employees. At the other extreme, most surveys in the

Stephen Cohen is Director of the Mathematical Sciences Research Center in the Office of Survey Methods Research, Dee McCarthy is an economist in the Office of Field Operations, Richard Rosen is an economist in the Office of Employment and Unemployment Statistics, and William Wiatrowski is Associate Commissioner in the Office of Compensation and Working Conditions, Bureau of Labor Statistics.

compensation and price programs are collected through a personal visit by a BLS staff member, with mail used for periodic updating of data.[2]

BLS has begun to give respondents to some of its programs the option of recording their responses on the Internet. The following programs either currently collect data via the BLS Internet facility or are testing such collection[3]:

- Annual Refiling Survey (part of the Quarterly Census of Employment and Wages)
- Current Employment Statistics Survey
- International Price Program
- Multiple Worksite Report Survey (part of the Quarterly Census of Employment and Wages)
- National Compensation Survey
- Occupational Employment Statistics Program
- Producer Price Index
- Survey of Occupational Injuries and Illnesses

Solicitation for Internet reporting varies from trying to entice reporters one at a time to become Internet reporters to giving all employers in the program the option of reporting on the Internet. This article focuses on two surveys that use different approaches—the Current Employment Statistics Survey (CES) and the Survey of Occupational Injuries and Illnesses (SOII). While both of these surveys have had employers reporting via the Internet for several years, and therefore have a significant body of experience, they provide contrasts as to type of survey and approach to solicitation. In 1996, the CES became the first BLS program to collect data via the Internet and solicits employers one at a time. The SOII, which began Internet collection in 2003, allows all respondents to use the Internet if desired and has the greatest number of Internet responses among all BLS programs.

## The CES and SOII programs

*CES.* The CES is a monthly survey of employment, payroll, and hours. The data collected are used to produce the establishment employment statistics contained in the *Employment Situation* release, a major economic indicator.[4] The sample of 400,000 business establishments provides data that are published after only 2½ weeks of collection. Detailed estimates of employment by industry, State, and area, average hourly earnings, and average weekly hours are published. Respondents are queried about seven basic items:

- All employees
- All employees payroll and commissions
- All employee hours (including overtime hours)
- Gross monthly earnings
- Production/nonsupervisory hours
- Production/nonsupervisory worker payroll and commissions
- Production/nonsupervisory worker hours (including overtime hours)
- Beginning in July 2006, CES will reintroduce collection of the number of employees that are women

The CES was originally a mail-in-mail-back design with data collection performed by BLS partners in cooperating State employment security agencies. Over the last dozen or so years, the CES has moved away from mail collection by exploring various modes that make it as easy as possible for a respondent to participate. The CES has used touchtone data entry (TDE) since 1986. This efficient and low cost collection mechanism is one of the primary methods of CES collection, accounting for 22 percent of monthly reporters. Respondents call a toll-free number and respond to a computerized interview. For large employers, the CES has arrangements to receive electronic files containing the required data elements. This collection protocol, known as electronic data interchange (EDI), is used for about one-third of CES reports each month.

The CES also operates several data collection centers, where reports are collected via Computer Assisted Telephone Interview (CATI) or via fax. Under CATI, an interviewer calls employers at an agreed-upon date and time each month to obtain their data. The interviewer enters data into an online system that performs a number of edit checks so that data can be directly validated with the respondent. Under fax collection, the BLS center generates a fax to the employer at the appropriate time each month. The respondent fills in the form and faxes the form back to the center for data entry. About one-fourth of CES responses are from CATI, while 13 percent are from fax. Newer approaches to data collection include fax optical character recognition and the Internet. In 1996, the CES became the first BLS program to capture data via the Internet, using a system developed exclusively for the CES program. In 2000, CES moved to the BLS-wide IDCF.

*SOII.* The SOII is a federally-mandated survey designed to yield the number and rate of workplace injuries and illnesses by detailed industry. The survey provides annual information on nonfatal injuries and illnesses occurring in the workplace, including worker demographics and case characteristics for those cases that require days away from work.[5]

The SOII is built on the collection of recordable cases—any occupational injury or illness that results in days away from work, loss of consciousness, restricted work or transfer to another job, or medical treatment (other than first aid). The Occupational Safety and Health Act requires that employers subject to the Act maintain a standard set of records covering their injury and illness experience. Each year, BLS

selects a sample of about 230,000 employers and asks them to submit reports on the summary of incidents at a specific employer location, with further sampling of cases to obtain details on the individual involved and the event, source, nature, and part of body affected.

Employers selected for the SOII are notified at the beginning of the year to maintain a log of their injury and illness experience over the entire year.[6] Data include establishment employment and hours worked, a count of injuries and illnesses, and, for those cases that involve at least 1 day away from work, some details on the individual and the incident. At the end of the year, employers provide these data to BLS, generally through individual State agencies.

The SOII is traditionally collected by mail. Internet collection was introduced in 2003 and has become a major source of data—accounting for more than 20 percent of the usable data on injury and illness cases. As is the case with the CES, the SOII continues to explore additional data collection modes. Beginning in 2005, the SOII introduced two additional electronic collection mechanisms—e-mail and electronic file transfer.

## Solicitation strategies

In setting up collection protocols, differences in survey design must be taken into account. The major difference in the design of the CES and SOII is that the CES is monthly and contacts the same employers for several years, while the SOII is annual and selects a new sample of employers each year. (Of course, many employers, typically larger ones, are selected for the SOII sample each year.) These differences lead to differences in the way Internet collection is solicited. CES targets individual employers for Internet reporting after first initiating them into the survey via the BLS data collection centers and collecting their data via CATI for several months. Because SOII is an annual survey, each employer is offered Internet reporting during the annual solicitation mailing.

CES. After several months of successful CATI collection of CES data, interviewers solicit respondents to participate in Internet collection. For those respondents who agree, the data collection center prints and mails an Internet package. The package includes a transfer letter with a temporary account number and password, a CES data collection form, and two brochures—*Reporting by Web* and *Establishing an Account on the IDCF*.

Beginning in the fall of 2003, the data collection centers began offering Internet reporting to respondents as one of the available self-reporting modes (TDE being the other primary self-reporting mode). Since then, about 200 employers per month, on average, have been converted to Internet collection. During this same period, about 1,000 establishments each month have been converted to TDE. As of early 2006, there are about 5,000 CES Internet reporters.

Nearly one-quarter of CES sample establishments self-report using their touchtone phone. These establishments would appear to be prime candidates for Internet reporting because they are already self-reporting. Because transitions can sometimes result in sample attrition, the CES program conducted a methodological study to determine the feasibility of and most effective way to transition touchtone establishments to Internet.[7]

Several alternative conversion/contact protocols were tested, including telephone, mail, and fax. The results of this study indicated that about 70 percent of current TDE units were eligible for Internet reporting (that is, they had Internet access and a compatible Web browser). (See table 1.) Of those that were eligible, about 90 percent elected to switch to Internet reporting. However, actual transition of these units to Internet reporting took some time, required considerable effort, and generally resulted in either neutral or reduced response rates. For example, after several months, only about 80 percent of the respondents that indicated they wanted to report by the Internet had activated their account. After intensive followup, response rates for the Internet group re-

**Table 1. Results from a test of alternative conversion and contact protocols—CES touchtone data entry to Internet**

| Unit | Total | | Contact mode | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Call | | Fax | | Mail | |
| | Number | Percent | Number | Percent | Number | Percent | Number | Percent |
| Sample | 3,011 | – | 1,002 | – | 1,001 | – | 1,008 | – |
| Contact rate (percent of sample) | 2,221 | 74 | 848 | 85 | 652 | 65 | 721 | 72 |
| Eligibility rate (percent of those contacted) | 1,583 | 71 | 476 | 56 | 543 | 83 | 564 | 78 |
| Conversion rate (percent of those eligible) | 1,412 | 89 | 413 | 87 | 493 | 91 | 506 | 90 |
| Activation rate (percent of those converted) | 1,084 | 77 | 309 | 75 | 408 | 83 | 367 | 73 |

turned to their preconversion rates when reporting by TDE. (See chart 1.) A number of respondents continued to report via TDE, even though they had indicated a desire to report by the Internet.

As a result of concerns that respondents were finding the system difficult to use or were not using it for monthly updating, the CES slowed the process of soliciting new Internet reporters in early 2005. Since then, BLS has been developing a streamlined Internet collection vehicle designed to accommodate the unique needs of the CES program—namely, just a few data items collected on a regular basis with a quick turnaround time. These new efforts are discussed at the end of this article.

*SOII.* The first Internet data collection for the SOII occurred in early 2003. In the initial SOII solicitation, data collection centers gave every potential respondent in 47 States a BLS Internet address to allow them to report their data via the Internet.[8] Beginning with 2004 data collection, respondents in all States received the option of Internet collection. In all years, the mailing to employers requesting survey data included a brochure explaining how to enter data using the Internet. Of the employers providing SOII data, 5 percent reported their data on the Internet in 2003. In 2004, the number of establishments supplying data via the Internet more than doubled to nearly 30,000 employers—about 12 percent of the sample. Employers also report details of injury and illness cases. By 2005, employers responding over the Internet reported more than 50,000 individual injury and illness cases—about 21 percent of the survey total.

All but a few of the nearly 30,000 SOII Internet responses from employers received in 2004 had sufficient usable data to generate establishment estimates. Establishments with 50 or more employees were more likely to report via the Internet than were smaller establishments. (See table 2.) These were also the establishments with the highest rate of injuries and illnesses. There is not a direct relationship between the percent of Internet responses and the rate of injuries and illnesses; however, the injury and illness rate tends to peak at 50–249 workers and then decline, while the percent who answer via the Internet gets larger as employment increases.

The SOII also captures details on individual injury or illness cases that result in days away from work. Overall, employers entered details of 52,044 injury and illness cases via the Internet in 2004. Interestingly, those with more cases (and thus more data to enter) tended to use the Internet system more than those with fewer cases. The average number

**Chart 1. Response rates: Current Employment Statistics touchtone data entry (TDE) to Internet conversion test, 2004**

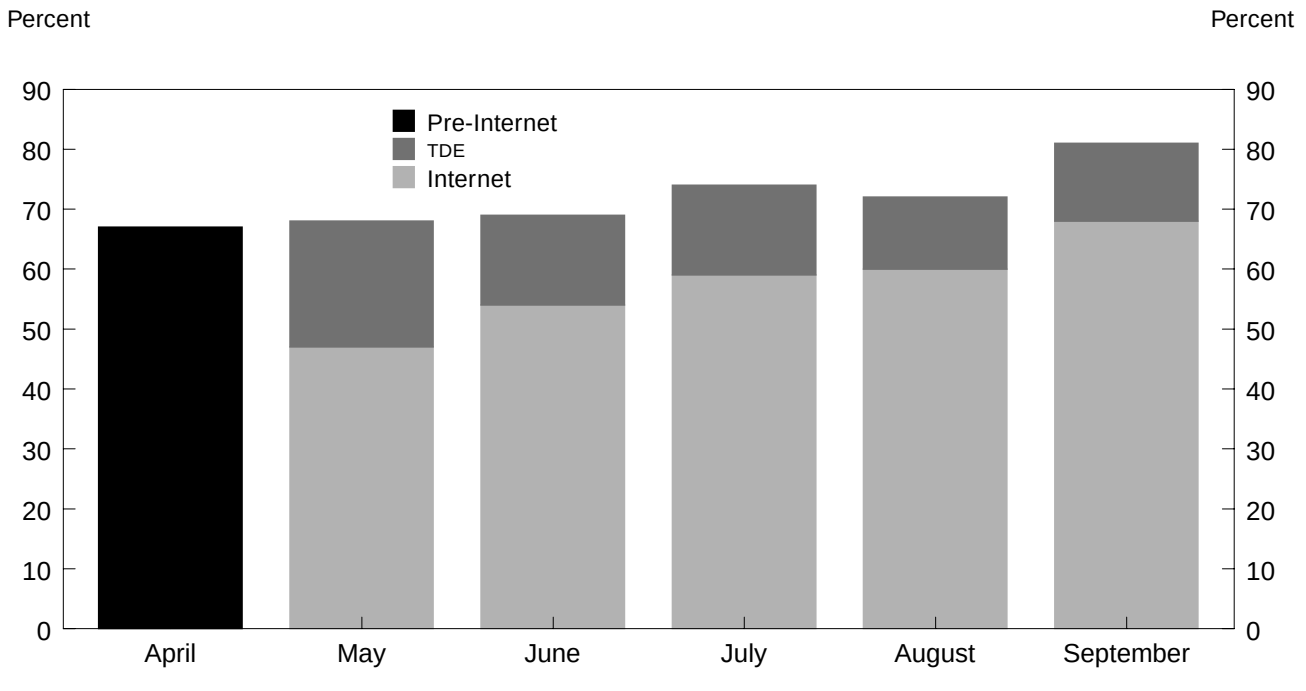


**Table 2. Survey of Occupational Injuries and Illnesses Internet responses and injury/illness rate by employment size, 2004 data collection**

| Employment | Percent of responses provided by Internet | Rate of injuries and illnesses[1] |
|---|---|---|
| All | 12 | 5.0 |
| 1–10 employees | 7 | 2.0 |
| 11–49 employees | 12 | 4.3 |
| 50–249 employees | 18 | 6.2 |
| 250–999 employees | 22 | 5.8 |
| 1,000 or more employees | 22 | 5.7 |

[1] Rate represents the number of injuries and illnesses per 100 full-time workers in 2003. The rates shown are for all responses, not just those provided by Internet.

of cases among Internet responders was 1.85, compared with 1.16 cases among mail responders. This may reflect the fact that a greater proportion of larger establishments reported data by Internet.

While Internet data reporting has a number of potential advantages, such as lower survey cost and improved data quality, BLS must continuously evaluate various operational issues to determine if Internet reporting will in the long run achieve these goals. For annual or one-time surveys, the prospect of reduced cost and improved data quality may be easier to realize because even if 10 or 20 percent of responses come in via the Internet, costs are reduced. For ongoing surveys that rely on quick turnaround, the operational issues of initial registration, timely response, nonresponse followup, and maintenance of contact information (such as the e-mail address) may make it more difficult to achieve the cost and quality objective. Tests designed to address these issues will be implemented in 2006, as described at the end of this article.

## Encouraging response

*CES.* Because the CES is part of the BLS monthly *Employment Situation* release, a respondent has a very short time to report data before the data must be processed for initial release. However, BLS is interested in obtaining as much data as possible for estimating labor force employment, and thus accepts data over several months before the estimates are final. Approximately 65 percent of the respondents who agree to respond via the Internet do so by the first deadline, the results of which appear in the monthly *Employment Situation* release. The response rate approaches 75 percent for the second deadline and is near 80 percent for the third and final deadline.

Under the Internet methodology, CES establishments that are expected to report by Internet receive e-mail messages prompting a response. Up to three e-mail messages are sent each month:

- Advance Notice: sent at the beginning of the reporting period
- Nonresponse Prompt: sent during the last few days of the reporting period
- Last Chance: sent the last day of the reporting period

The text of each message is somewhat different and geared toward the specific point in the collection period. (See box for sample messages.)

In general, about 50 percent of the sample will report after receiving the initial Advance Notice message. After the Nonresponse Prompt and Last Chance messages, an additional 15 percent report, making up the 65 percent of initial respondents whose information appears in the *Employment Situation* release.

Internet response rates are not as high as the rates achieved

### Sample of Current Employment Statistics e-mail contact messages

**Advance Notice message:**

The data reported to the Current Employment Statistics program last month were included in the estimates reported by the Bureau of Labor Statistics on Friday, February 06, 2004, and publicized by major media organizations such as The Wall Street Journal.

Please report your February employment, payroll, and hours data by Friday, February 27, 2004, if possible. Our Internet address is https://idcf.bls.gov. If you have any questions, please contact cesmail@bls.gov.

**Nonresponse Prompt message:**

As of today, our records show that we have not received your employment, payroll, and hours data for February. Please report these figures by Friday, February 27, 2004, if possible. If you have reported your information by some method other than Internet, please send a note to cesmail@bls.gov. Please report your data to https://idcf.bls.gov. If you have any questions, please contact cesmail@bls.gov.

**Last Chance message:**

Urgent Request for Data: This is your last chance to report your February 2004 employment, payroll, and hours data in order for it to be included in our preliminary estimates of February employment. The Bureau of Labor Statistics will publish the February Employment Situation report on Friday, March 05, 2004. The Employment Situation report is a monthly news release that highlights the previous month's employment estimates. Your figures help to make these estimates possible. Please report your data to https://idcf.bls.gov.

for establishment data collected by TDE, where the response is 84 percent by the third and final deadline. The reasons for these differences are not completely clear. The relative characteristics of the two populations are essentially the same—generally, employers that have been successfully reporting via CATI for several months and who report for one or only a few locations. There is some evidence that the additional steps required for initial registration on the BLS Internet site and other related password/browser problems that are not present with the touchtone system are factors in the response difference. In contrast to the Internet, TDE does not require respondents to have a separate account/password, and the system works from any touchtone phone.

Because virtually all contact with Internet establishments is by e-mail, maintenance of the e-mail address is critical. The Internet page provides a utility for the respondent to update their e-mail address online. These updates can be extracted and transferred to the main database where the CES sample of employers is maintained. CES staff closely monitors the e-mail delivery rate and follows up on undelivered e-mail by telephone. A recent review shows that e-mail delivery rates are about 97 percent. Establishments reporting by TDE generally receive their notification by fax. Delivery rates by fax similarly are about 96 percent.

*SOII*. Each year, employers in the SOII receive a flyer providing instructions on how to enter the data via the Internet. The SOII collection period extends from approximately January to July of each year. Nonresponders receive two additional mailings to encourage response, including Internet response. Respondents who begin entering their SOII data via the Internet but do not complete the survey are contacted by e-mail prior to the end of data collection. The following is a typical message:

> Thank you for visiting the BLS Internet Data Collection Facility recently to report your ***2003 Survey of Occupational Injuries and Illnesses*** data to us. We have completed a review of our database to identify employers that registered with our website during the time period of {dates} but have not yet completed and submitted their survey response.
>
> We do not have a completed survey response from you and your data are very important to us.
>
> If you have returned your survey by mail, please e-mail **osh.helpdesk@bls.gov** to let us know. If you intend to respond to our 2003 Survey of Occupational Injuries and Illnesses using the BLS Internet Data Collection Facility rather than mailing your completed survey request, then please return to the website at **https://idcf.bls.gov** to complete and submit your response.
>
> If you are experiencing problems with our website or with submitting your completed response, please contact us at **osh.helpdesk@bls.gov**. A staff member will contact you within 24 hours Monday–Friday by return e-mail to aid you in any way we can.

In some cases, respondents did not realize that they had not completed the survey. The 2004 Internet collection incorporated changes to the screen design and instructions to help alleviate this problem. For the 2005 collection, a further revision to the SOII Internet collection screens allowed respondents to enter changes even after they had submitted their complete responses.

## Data security

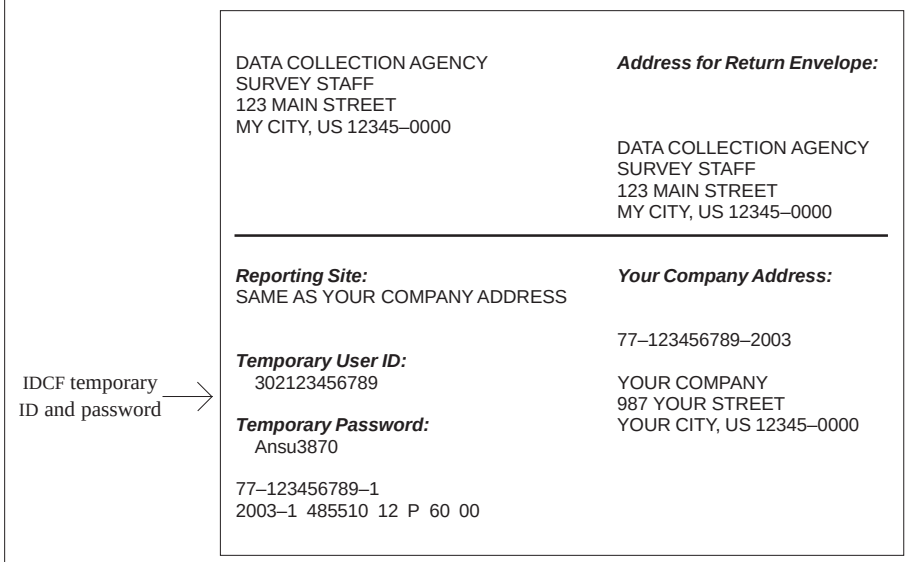
One concern that BLS has with Internet data collection is data security, including data interception during transmission and unauthorized users corrupting data on a BLS computer. To guard against unauthorized activities, the IDCF imposes strict controls on respondent authentication.

Respondents initially log in using the temporary account and password located in their solicitation material. (See exhibit 1.) Users then receive a permanent account and password, similar to that used for Internet access to personal financial institution data. Accounts and passwords provide flexibility for the respondent, because they can use almost any PC connected to the Internet to report their data. Such flexibility also entails some risk, because accounts and passwords may be stolen or forgotten. BLS does not require respondents to update their passwords, which eases the burden on them, but can result in some increase in security risk.

Users who require a greater level of security can invoke the option to obtain a digital certificate—a small file on the client's PC that authenticates the respondent's identity. The digital certificate has the following properties:

- Embedded in client browser
- Uses Public Key Infrastructure (PKI)[9]
- Ensures identity of respondent to BLS
- Ensures identity of BLS to respondent
- Eliminates the need for respondent to enter account and password information
- Gives ability to send and receive secure e-mail

**Exhibit 1.** **Example of mailing label from Survey of Occupational Injuries and Illnesses**




## Help desk

Both the CES and SOII programs maintain a "help desk" for employers that have reporting problems. In addition, there is a central IDCF help desk that is intended to respond to more global questions, often related to system access. As respondents report via the Internet, issues can surface. Most issues can be categorized as follows:

- Account setup and initialization
- Login problems
- Data problems

*Account setup and initialization*. Because all first-time users must access the system using a "temporary account/password" to initialize their account, these users may experience difficulties with initial startup. For example, for security reasons, the temporary password contains a random string of case sensitive letters and numbers. Respondents sometimes have difficulty reading/entering these digits correctly. Another issue may be the respondent's browser, as some browsers are not compatible with the BLS Internet system.

*Login problems*. Respondents may have difficulty logging in because they have forgotten or misplaced their login id or password or have entered them incorrectly. Also, changes in computer hardware or configuration may cause a digital certificate to fail. A related issue here is changes in the respondent contact person. If the existing contact leaves the company, the prior contact may not transfer the account/password information to the new contact. If the new contact is using a digital certificate and has a different PC, the certificate must be transferred to that PC or they must obtain a new certificate. Among CES respondents, of those who use digital certificates, about 4 percent have requested help with the digital certificate. Likewise, of those who use an account/password, about 4 percent have requested help with the account/password. About one in four help requests for those attempting to use the Internet to enter SOII data are for help in entering the system, such as a lost account/password or problems with the digital certificate.

*Data problems*. Data problems occur when the respondent has questions about the data elements themselves. While each survey provides a variety of context-sensitive help menus, respondents may still have questions that need to be

Digital certificates offer a higher level of security than passwords but give the respondent less flexibility. With a digital certificate, the respondent must use the PC on which the certificate resides. They provide the highest level of security in terms of data transmission. In addition, because the certificate resides on the respondent's PC, there is a very low risk of it becoming lost or stolen. (Of course, if the certificate resided on a laptop computer, there is a greater risk of the laptop being lost or stolen.) Because BLS assumes the cost of obtaining and maintaining the certificate, the use of the certificate does increase the cost of survey operation. Generally, certificates must be renewed every 2 years, or when a new PC or operating system is installed.

Prior to July 2005, respondents were required to choose between a digital certificate and an account/password before they could proceed to enter data. There is some evidence that this choice was confusing to respondents. This experience, and an analysis of help desk requests on this issue, led BLS to change its protocol, making the account/password the default option. Most respondents currently use the account/password option.

Regardless of which security option the respondent selects, all data transmitted via the BLS Internet facility is encrypted using a 128-bit secure sockets layer at the browser level. A secure sockets layer is a security protocol that provides communications privacy over the Internet. The protocol allows client/server applications to communicate in a way that is designed to prevent eavesdropping, tampering, or message forgery.

addressed. For the SOII, State employees, working with respondents in their States, typically handle these types of questions among mail respondents. The move to Internet collection has shifted many of those questions from telephone to e-mail and has shifted responsibility to those handling the IDCF help desk. About one-quarter of questions on the SOII survey were requests to change data after a respondent had completed the survey. In 2003 and 2004, respondents were not able to make changes after entering their Internet submission. BLS and State staff had to take the revised data and enter it into the BLS processing system. In response to this issue, BLS changed the 2005 system so that respondents now have the opportunity to make data changes even after their data have been submitted.

The CES program recently completed a comprehensive review of help desk inquiries in an effort to measure both the volume and type of contact. Help desk inquiries have been running at about 5 percent of the active establishments. (See chart 2.) The large spike in January 2004 is a function of the annual mail out of new collection forms. This generally triggers a large number of questions. The smaller spike in May 2004 reflects the start of the TDE to Internet conversion test when CES staff added more than 1,000 new Internet establishments.

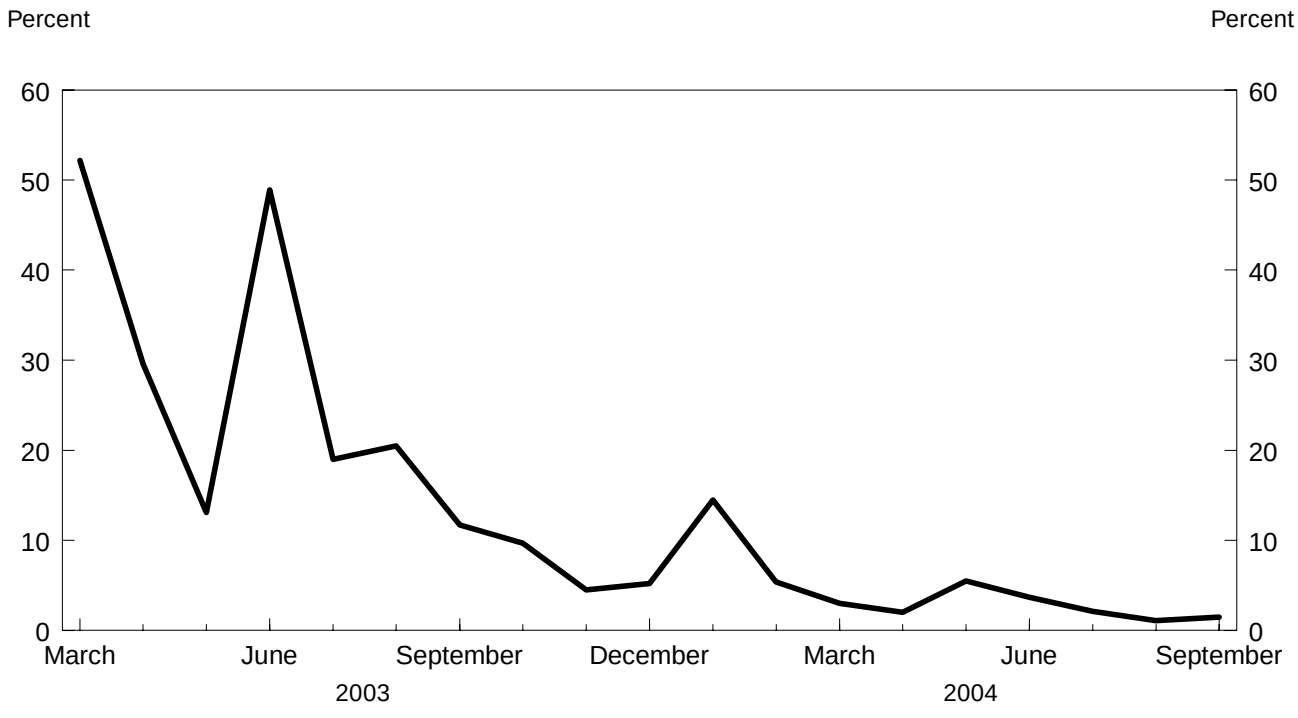
## Data quality

An advantage of Internet collection is online editing—the visual interface allows interaction with respondents that is not available in other self-administered collection methods such as fax or mail. BLS has experimented with a variety of alternative editing techniques for Internet collection, attempting to balance the desire to obtain high-quality data with the need to limit the burden placed on respondents.

For example, under TDE where the respondent simply receives a set of prerecorded prompts, it would be very difficult to try to reconcile data that fail an edit. The respondent would have to listen to a long question, and then select a number of options to correct one or more data elements. However, when reporting online, the respondent sees a replica of their report form and can be provided more instructions on the nature of any edit question and directions on how to correct the error (if necessary).

Currently, the BLS Internet Data Collection Facility uses two alternative approaches to editing—individual screen editing and one-time editing. In SOII, the edits occur when the respondent presses CONTINUE to go from one screen to the next. Instead of moving to the next screen, the current screen is refreshed with a red bold indicator showing

**Chart 2. Help desk inquiries as a percent of active Internet establishments, Current Employment Statistics, 2003–04**



that an error exists. The screen prompts respondents to change data or provide an explanation of unusual data. In the CES, which has fewer data elements than SOII, edits appear after a respondent has entered all of the data elements and selected the SUBMIT button. If there is a failure, an edit box appears. Edit messages appear as highlighted text below the respondent's data. Respondents have to reconcile the error or enter an appropriate comment code to complete their submission.

Both surveys use mandatory and optional edits. For mandatory edits (for example, the number of women workers cannot exceed total employment), the respondent must reenter correct data. For optional edits, the respondent may leave the existing data and provide a comment code or explanation indicating that the data are valid.

*CES.* The initial CES Internet system performed a few basic edits, such as:

- Logic errors—All Employment must be greater than or equal to Production Workers
- Range checks—average hourly earnings between $1 and $150
- Validity checks—numeric entry, mandatory fields completed

The results from these initial basic edits were as follows:

- Over a 12-month period, approximately 40 percent of Internet establishments failed at least one edit check
- Approximately 3 percent of Internet establishments failed one edit check each month
- In 88 percent of all edit failures, the respondent corrected and resubmitted data during the same session

Because CES data can be quite variable from month to month, most edits are optional. Experience with review of edits tends to suggest that most records that fail current machine edits are acceptable and ultimately used in estimation. Therefore, the only mandatory edits relate to impossible data situations or entry of nonnumeric values in the data fields.

Enhancements to the current CES Internet application in May 2001 included 21 edits. These additional edits expanded on the existing logic-type edits and included:

- Industry-specific tolerances for average hourly earnings and average weekly earnings (minimum and maximum values)
- Over-the-month change values to compare each data element against broad industry norms

BLS patterned the enhanced edits after those currently used by CES in its CATI system.

With the introduction of enhanced edits, the incidence of edit failures rose from 3 percent of reports each month to about 7 percent. This was to be expected because many additional edit checks were added. The percent of establishments entering a comment code likewise rose from about 6 percent to about 14 percent. By comparison, the proportion of touchtone self-reported establishments that enter a comment code is 5 percent. During CATI interviews, the percent of records with a comment code is about 12 percent. Thus, it appears that Internet reporters are self-reporting comment codes to explain large fluctuations in the data at about the same rate as interviewers are entering them, and considerably more often than touchtone respondents self-report a comment code.

After introduction of the new edits, CES staff conducted several debriefing calls to respondents that had triggered one or more of the enhanced edits. Respondents did not express any concerns about the edits and were able to navigate through the process either by correcting the data or providing a comment code.

*SOII.* With the first release of an online Internet instrument for the 2003 collection, SOII did not want to overwhelm respondents with too many error messages. There was a conscious decision to keep the editing to a minimum—flagging only invalid data that would cause a database problem—so as to get the respondents in and out quickly. The edits included in the Internet instrument ensure validity of the data; for example, the values entered in Annual Average Employment and in the Total Number of Cases with Days Away from Work must be numeric.

In the SOII production data collection systems, there are nearly 170 edits. The edits fall into these categories:

- Validation edits (data are in the correct format)
- Consistency (relationships between data elements are satisfied, for example, total hours must be in the range for the economic activity and employment level of the establishment)
- Reasonableness (primarily for coding consistency—can you really have a funeral director working in a furniture store?)

The SOII production system edits establishment data for differences in total reported employment from the original sample file, data out of range, and injury and illness counts unusual for the reported industry. The information requested for individual injuries or illnesses is edited for consistency between the nature and source of the injury versus the injury outcome. For example, if the nature of an injury is a sprain, the body part affected cannot be the brain or skull.

State employees review and correct these data. In cases where employer-provided data cannot be easily reconciled, State employees will call respondents for clarification. This is true whether the data were provided by mail or via the Internet.

For the 2003 SOII data collection, 50 percent of employer responses had edit failures, and 45 percent of the cases had error messages. In many cases, there were multiple errors from the same employer. The edit failure rates for responses received via the Internet were no different than those received by mail collection. Less than 1 percent of Internet responses could not be used because errors could not be resolved; this rate is identical to that of mail responses.

BLS has embarked on developing a research program on the cognitive aspects of edits—how they are shown, what wording is used, and what options are presented. Built-in edits or data checks are potentially an important tool for improving the quality of data. Because respondents can correct survey entries, a reasonable assumption is that online edits will produce higher quality data. In addition, online edits may result in other efficiencies, such as lower survey costs, because fewer followup calls should be necessary.

Past testing of IDCF applications, however, has revealed two usability problems with the implementation of some edit messages. First, users sometimes completely miss seeing the edit message. Second, even if users do see and read the message, they may not understand the message or the action that should be taken. In addition, there are other issues associated with the use of edits. Although edits are potentially useful for improving data quality, if overused, poorly designed, or confusing, they might increase respondent burden significantly and, therefore, have negative impacts on survey response or even data quality. With these considerations in mind, the primary purpose of ongoing research is to identify some key design features that lead to edit messages being noticed and errors being correctly resolved.

## New activities in 2006

BLS has introduced Internet-based collection into its lexicon of data collection tools, but must continue to look for ways to improve its methods. In the case of the CES program, which operates under a tight monthly collection time frame and where participation is voluntary, Internet collection must be as easy as possible for employers to use, with the goal of attracting more employers to use this mode of collection each month. For SOII, where a significant portion of all data are now captured via the Internet each year, there is interest in moving away from paper questionnaires and making the Internet the standard collection mode. In pursuit of these goals, both programs are undergoing Internet collection tests during 2006.

The CES Internet test involves the development of a new, simplified Internet vehicle for employers to use when entering their data. Unlike the existing Internet screens that are standard in the BLS Internet Data Collection Facility, the CES screens would have a simplified access mechanism where the user would enter their survey identification number and automatically be linked to a blank data entry screen.[10] Employers will simply enter their current data and click on the "submit" button. Unlike the system used for Internet collection for other BLS programs, there would be no ability to save data and return at a later time, nor would the user have access to prior period data. The intent is to provide a quick and noncumbersome method of one-time data entry, designed to facilitate collection of the small number of data elements in a short time period. This experimental system is expected to be operational for a test of about 100 establishments in early 2006.

In addition, the CES program is testing an e-mail collection system. Respondents will be sent a replica of the CES collection form via e-mail. Data are entered directly on the e-mail form, after which the respondent clicks on the SUBMIT button to transmit data to BLS.

For SOII, Internet response has been so positive that there is an interest in eliminating the massive printing and mailing operation that accompanies paper questionnaires. The long-run goal is to provide employers with a brief notice (perhaps a letter or a postcard) indicating that it is time to enter their occupational injury and illness data into the Internet system. A prototype for such a process is being tested with a small number of employers during 2006 data collection. Three groups of 2000 employers, plus a control group (each a small fraction of the approximately 200,000 employers from which data will be collected), will test new collection protocols during 2006.

Test group A will receive a pamphlet indicating that they can provide data by Internet or e-mail, or can receive a paper form by calling BLS. Group B will receive a similar pamphlet indicating that they can provide data by Internet or e-mail, or can call BLS for assistance. Both groups A and B will receive reminder mailings if they do not respond within specified times. Group C will receive the same information as group B, except that their reminder mailing will include a paper questionnaire. The three samples plus the control sample are made up of establishments in similar industries and geographic areas, so that the results can reveal whether these different solicitation treatments had any effect on response, method of collection, and timing of response. Analysis of the test data will help the SOII program determine if, in the future, it can expand the number of establishments that do not receive paper questionnaires.

These tests are just some of the types of research that BLS will continue to conduct on its Internet collection activi-

ties. By learning more about differences in collection response due to different survey protocols and timing, BLS may be able to make more efficient use of Internet collection in the future. □

## Notes

[1] In addition to collecting data from employers, BLS also conducts surveys that collect data from individuals. One such survey, the Consumer Expenditure Survey, is currently undertaking a pilot test of Internet data collection from individuals.

[2] BLS conducts two types of programs that collect data from employers. Federal-State cooperative programs are conducted in conjunction with the States; typically State employees are responsible for data collection. Directly-collected programs do not have a State component; all data are typically collected by BLS employees. In most programs, data collection involves an initial contact for collection of detailed data (referred to as "initiation" collection) followed by brief periodic contacts for collection of the most current data (referred to as "update" collection or "repricing"). Employer data are generally collected for a single location, such as a store, plant, or central office. Such locations are referred to as establishments. This article will refer to employers, establishments, and respondents. In each case, the reference is to a single location and the data from that location.

[3] Information on these BLS programs is available on the BLS Internet site at **www.bls.gov**.

[4] The *Employment Situation* is a monthly news release that contains information on current employment and unemployment statistics. More information on the Current Employment Statistics program is on the Internet at **http://www.bls.gov/ces/home.htm**.

[5] More information on the Survey of Occupational Injuries and Illnesses is on the Internet at **www.bls.gov/iif**.

[6] Some employers are required by the Occupational Safety and Health Administration to maintain a record of injuries and illnesses. Employers who do not have such a requirement may be selected to be included in the BLS survey. In such a case, those employers must maintain a record of injuries and illnesses for the year that they are in the survey.

[7] Richard Rosen and Tony Gomes, "Converting CES Reporters From TDE to Web Data Collection," presented at the American Statistical Association Conference, August 2004.

[8] Some slight differences in the collection form and requirements in three States made it impossible to offer Internet collection in 2003. This was resolved prior to the start of the 2004 collection.

[9] In the United States, the National Institute of Standards and Technology (NIST) is taking a leadership role in the development of a Federal Public Key Infrastructure that supports digital signatures and other public key-enabled security services. NIST is coordinating with industry and technical groups developing PKI technology to foster interoperability of PKI products and projects. Further information can be found on the NIST Internet site at **http://csrc.nist.gov/pki/**.

[10] A visual verification system, where the user has to enter a designated string of characters, will be used to prevent automated computer programs from sending multiple responses to the CES Internet system. For visually-impaired users, an override of the visual verification system will be available.