

Encouraging the Use of Alternative Modes of Electronic Data Collection: Results of Two Field Studies June 2007

Kathy Downey, Dee McCarthy, and William McCarthy¹
U.S. Bureau of Labor Statistics¹

Abstract¹

In an attempt to lower postage, processing costs, and printing costs, as well as increase Internet usage, BLS performed two different data collection tests. The tests were conducted during 2005 and 2006 for the Survey of Occupational Injuries and Illnesses. For these tests, BLS did not mail the usual 12-page survey booklet to the establishments in the test groups. Instead, a 4-page booklet that described how to report via the Internet or email was mailed (“electronic options booklet”).

Results showed that the Internet response rate increased with the mailing of the reporting options booklet instead of the standard 12-page survey booklet. Also, data processing time decreased when respondents used the Internet or e-mail for reporting. However, the overall response rate decreased when respondents used electronic options for reporting data. Further field studies are planned to explore these results.

Keywords: web survey, response rates, e-mail survey, data collection

1. Introduction

The BLS was interested in expanding the use of Internet data collection for the SOII because it reduces processing time without increasing data input for the BLS, the States, or the contractors used for this survey. Also, a benefit of Internet data collection is the capturing of narrative text describing days-away-from-work cases, which in turn allows more detailed analysis of the characteristics of workplace incidents. Currently, the details for days-away-from-work cases are sometimes not data entered in the SOII database.

In addition, at the direction of the Congress, the BLS is moving to new methods of accounting for mail costs. This has led to an examination of the number of pieces of mail the SOII sends, including weight and postage. The data collection test documents for the SOII were designed to weigh less and therefore incur less postage.

¹ Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

In an attempt to lower postage and processing costs, lower printing costs, and increase Internet usage, the BLS performed two different data collection tests for the 2005 and 2006 surveys.

1.1 The Survey of Occupational Injuries and Illnesses (SOII)

The Bureau of Labor Statistics (BLS) conducts the federally-mandated Survey of Occupational Injuries and Illnesses (SOII). The SOII is a Federal/State cooperative survey that collects data from a sample of approximately 230,000 establishments each year². The data collected are based on the Occupational Safety and Health Administration’s recordkeeping requirements (29 CFR Part 1904 and 1952). The survey collects data on workplace injuries and illnesses including total recordable cases, cases with days away from work, cases with days of job transfer or restriction, and other recordable cases. Respondents are directed to record the demographic and case characteristics of the more serious cases that involve days away from work. The demographic characteristics collected are race or ethnic background, age, length of service with the employer, gender, and occupation. Case characteristics are time the employee began work, time of the injury or illness, the nature of the injury or illness, the body part affected, the source of the injury or illness, and the event or exposure that resulted in the injury or illness. The survey is mandatory.

1.2 BLS Internet Data Collection Facility

Respondents to the SOII were first able to use the Internet Data Collection Facility (IDCF) in the 2002 survey year. The IDCF is the BLS’ centralized data collection facility used by SOII and other BLS programs as the platform for Internet data collection. The IDCF provides a uniform, manageable, and secure architecture for Bureau surveys to collect information

² Sample establishments are selected from the unemployment insurance files maintained by BLS. The establishments include private industry and State and local government, with industry defined by the North American Industry Classification System (NAICS).

over the Internet.

When using the IDCF to report for SOII, employers can enter their injury and illness data along with their employment and hours worked. The online, Web form is designed to look similar to the mailed form. For the 2004 SOII, 29,551 establishments used the IDCF to submit data. These establishments also submitted 50,707 days-away-from-work cases (a separate sub-form is completed for each case).

1.3 E-mail Reporting Option

In the 2004 survey year, BLS offered e-mail data collection for the first time. Employers desiring to use this option sent an e-mail to a specific BLS e-mail address. They automatically received a Microsoft WORD template of the survey form by return e-mail. Employers entered their data and then e-mailed the template as an attachment to a State-specific, BLS e-mail address. State personnel were then required to process the e-mail survey form in the same manner as hard copy survey forms returned by employers via the mail. In 2004, 914 establishments reported their SOII data using e-mail.

2. Methods

2.1 2005 Survey Test

For the 2005 survey test, there were three treatment groups and one control group. All of these groups were of equal size, 2,000 units each, but differed in their method of data collection and survey materials.

2.1.1 Sampling

Units that were collected on the Internet in survey year 2004 were removed from consideration. Also, based on other exceptions³ the entire 2005 sample was not in-scope for these tests. Thus, no national estimates could be computed using the results of this study.

Each data collection test sample included 2,000 units, for a total of 8,000 units in the study. To ensure that the samples were homogeneous, four units were drawn from 2,000 randomly selected State sampling cells. A sampling cell was defined as a

³ These exceptions included all of Puerto Rico, Guam and the Virgin Islands; units where the employer was expected to sub-sample their days-away-from-work cases; sample units where the special handling field was not blank; and Maine, New Jersey, Illinois, and Wisconsin public sector units.

State/ownership/industry/size class group. One unit from each set of four was randomly assigned to either one of the test groups or to the control group. In designating the units for each of the four groups, only sampling cells with four or more sample units were considered.

The number of cells used from an individual State was proportional to the overall sample of the State that remained after the units discussed above were removed. This meant that States with larger overall total samples were selected for the tests at a higher rate than States with relatively small samples.

2.1.2 Survey materials

The mail-out envelope for the treatment groups contained a 4-page document giving respondents the following two reporting options:

- 1) Use the BLS IDCF system; or
- 2) Send an e-mail to the BLS address and receive an automatic reply with a copy of the survey attached. The return e-mail message contained a list of State-specific e-mail addresses so that the respondent could e-mail the completed booklet to the appropriate State office.

Test Group 1 received the 4-page document that includes the options listed above in all three mailings (first mailing, and first and second nonresponse mailings). In addition, their 4-page document stated they could call a telephone number and request a 12–page hardcopy of the survey. (It was explicitly stated they could request a hardcopy of the survey.) This telephone number was answered by employees in the BLS national office.

Test Group 2 received the 4-page document that includes the options listed above in all three mailings (first mailing, and first and second nonresponse mailings). Test Group 2 received the same materials as Group 1, except their materials did not explicitly mention the option of getting a hardcopy of the survey and neither did the people answering the help number. If the respondents phoned the help number, the national office personnel did not initially volunteer the information that a hardcopy of the survey was available. If the respondent was unable to use the Internet or called to refuse to participate in the survey, BLS encouraged the respondent to use an alternative approach (e-mail them a copy of the survey, have them print it, complete it, and mail it to the appropriate State office). If the e-mail option was not acceptable to the respondent, BLS then volunteered information about a hardcopy survey and mailed the respondent a survey

booklet that included a business-reply envelope.

Test Group 3 is identical to Test 2 for the first mailing and the first nonresponse mailing, including the procedure that the hardcopy of the survey was not mentioned as a reporting option. However, for the second non-response mailing, respondents were mailed the standard (control) 12-page survey booklet and a 2-page document describing electronic reporting options, along with the non-response letter and the non-response return envelope.

NOTE: In survey year 2004, there were 6,688 units collected by the Internet collection method that were also sampled for survey year 2005. These units are termed “INET” units in this study. All of these INET units, except special handling units and units where the employer was expected to sub-sample their cases, received the Test 2 option. However, these INET units are not included in the Test 2 group in the analysis of the tests. Because they were previous respondents via the Internet, the BLS felt they would be more likely to respond in the same manner in survey year 2005.

The control group received the standard 12-page survey, along with a 2-page document describing electronic reporting options in all three mailings (first mailing, first and second nonresponse mailings).

The first mailing occurred in January 2006. The first non-response mailing was February 21st and April 5th was the second non-response mailing. On May 9th, the sample was handed over to the States for their non-response efforts. This marked the end of BLS involvement in data collection and is used as the closing date for the 2005 test. Data collection for the 2005 SOII closed on June 26, 2006.

2.1.3 Analyses

Two general types of measures used in the analysis of the SOII study: 1) measurements of response rates and 2) estimates of data collection burden, as defined as respondents requesting help from BLS or the States, and data processing time. For response rate measurements, rates for total response and Internet response were computed and compared to the entire SOII sample. The response rates were calculated as the number of respondents divided by the number of eligible establishments (excluding out-of-business, out-of-scope, and duplicates). The Internet response rate was the number of Internet responses divided by eligibles.

For data collection burden, BLS analyzed the percentage of units contacting the Help phone number and processing time and costs.

2.2 2006 Survey Test

2.2.1 Sampling

The 2005 survey year test compared results using employers from the same sampling strata – industry / size class. However, the 2005 test was limited to 2,000 sampling strata nationwide. With the possible implementation of electronic collection for all sample units beginning in survey year 2007, the BLS needed to determine if the same type of results would be seen when all sample units in a State were given the electronic collection option. Therefore, BLS needed to replicate the 2005 test, but change the sampling plan so that whole States were selected for the test. However, rather than compare three treatment groups, as was done in the 2005 survey year, in the 2006 survey year Group 2 was the only treatment group.

The States selected for the 2006 test represent two general groups: those that are State partners⁴ (California, District of Columbia, Florida, Illinois, Maryland, New York, North Carolina, and Virginia) and States collected by the BLS regional offices (Colorado, Idaho, Mississippi, New Hampshire, North Dakota, Ohio, Pennsylvania, and South Dakota). A total of 16 States had all of their SOII sample participate in the test.

Another group of employers of interest in this test was the INET units. These were employers who responded by the IDCF in the 2005 survey year and were also selected in the 2006 survey year sample. There were 6,688 INET units in the 2005 survey year, but due to the success of the 2005 survey year tests in increasing the usage of the IDCF, the number of INET units for the 2006 survey year rose to 13,297. The total number of units in the test for survey year 2006 was 88,421.

2.2.2 Survey materials

As mentioned previously, those units in the treatment group used the Test Group 2 methodology as described in the 2005 test. That is, they received the 4-page form outlining the two reporting options, Internet or e-mail, and when respondents phoned the help number, the State personnel did not initially volunteer the information that a hardcopy survey form was a reporting option.

⁴ These States collect the data themselves and then send the data files to BLS.

The first mailing was sent out January 4, 2007. The two nonresponse mailings occurred March 2, 2007 and April 20, 2007. Data collection close-out was July 17, 2007.

2.2.3 Analyses

The 2006 test used the same analyses as the 2005 test. However, the analyses were not split by the end of the test period (when the sample is handed over to the States) and the end of the data collection (the final close-out). Only the latter numbers are presented.

3. Results

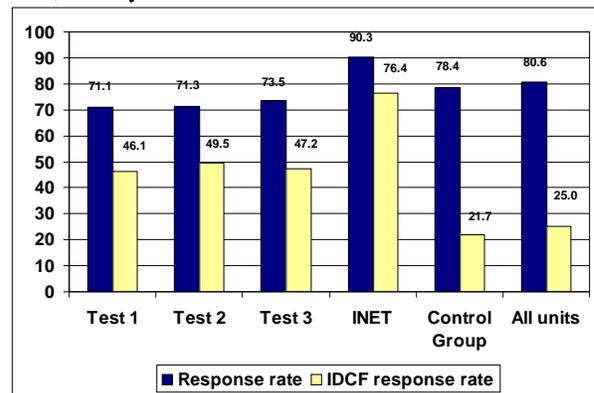
3.1 2005 Survey Test Results

3.1.1. Response rates

As shown in Figure 1, the group with the highest response rate by the end of the test period⁵ was the INET group (90 percent). This stands to reason because the INET establishments who used the IDCF the year before are very willing to use it again and may use it sooner in the collection cycle.

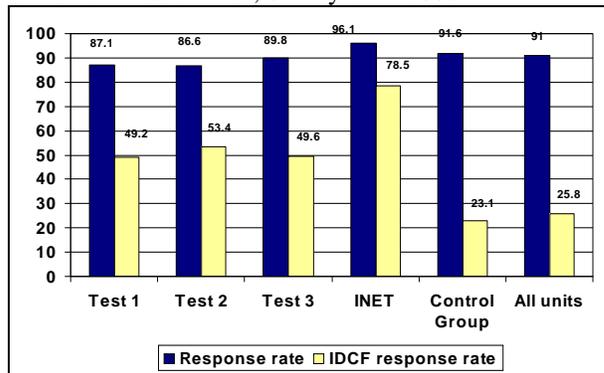
The control group obtained the next highest response rate at 78 percent. The three treatment groups' response rates ranged from 71 (Test Groups 1 and 2) to 74 percent (Test Group 3). When we examined the response rate over time, Test Group 3's rate pulled away from the other test groups' rates after the second nonresponse mailing, which is when Test Group 3 non-respondents were automatically mailed the hardcopy of the survey. This suggests that some employers were unwilling, or unable, to report using electronic options.

Figure 1. Response Rates by Study Group, at End of Test, Survey Year 2005



The same trend continued until the closing of the data collection, as shown in Figure 2. The INET group had the highest response rate (96 percent), followed by the control group (92 percent) and the test groups (87 to 90 percent). That means that the test groups' response rate lagged behind the control group's by two to five points at the end of data collection.

Figure 2. Total Response Rates by Study Group, at End of Data Collection, Survey Year 2005



All three test groups far outperformed the control in terms of the proportion of employers who responded using the Internet (49 to 53 percent for test groups versus 23 percent for the control group; Figure 2). This could partially be explained by a lack of larger establishments in the control group. There were no large employers (1,000-plus employees) and only five employers in the next largest size class (250 to 999 employees) in the control group. It is thought that large employers are probably more familiar with and have better access to the Internet than the smaller employers. The percentage of establishments using the Internet varied by establishment size and ranged from 20 percent for the smallest size class (1 to 10 employees) to 31 percent for the largest size class (1,000-plus employees).

⁵ Again, this is the time the sample was handed over to the States for their non-response efforts, not the final closing of data collection.

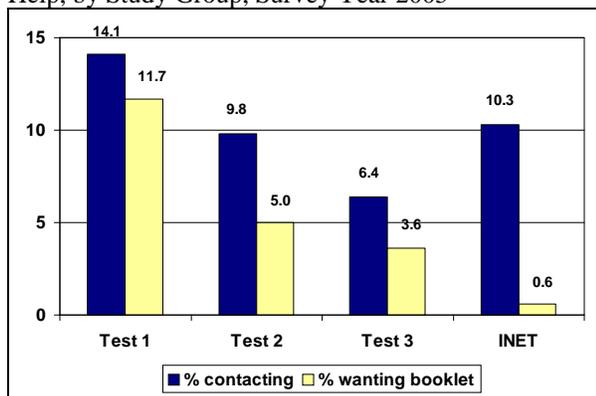
There were no distinguishable trends for response rates by industry. There were also none by State.

In summary, all three test groups ended with an Internet response rate around 50 percent. IDCF usage results in significant savings in data entry costs, as well as mailing costs and printing costs. More information on these data entry savings will be shown in this report in the section on the 2006 SOII tests.

3.1.2. Establishments needing help

Test Group 1 establishments contacted the Help phone number at a greater rate (14 percent; Figure 3) than any other test group. And as would be expected, this group requested a 12-page survey booklet at a much higher rate than the other test groups since it was given as an option in the survey materials. Approximately five percent of Test Group 2 and four percent of Test Group 3 employers requested a 12-page survey booklet.

Figure 3. Percentage of Establishments Calling for Help, by Study Group, Survey Year 2005



Most of the calls from the three test groups were to request a hardcopy of the survey. Help with the IDCF was the second most reported reason. Calls from employers in the INET group were primarily for help with the IDCF. Most often, these were related to access⁶.

Were the 2006 tests for survey year 2005 successful in raising Internet and e-mail response? With Internet submissions rising by 81 percent (from 29,551 establishments in survey year 2004 to 53,575 establishments in survey year 2005) and e-mail collection rising by 171 percent (from 914 in survey

⁶ To address these concerns, BLS implemented improvements to IDCF access for the next survey year (survey year 2006).

year 2004 to 2,475 in survey year 2005), the goal of increasing the use of the electronic options for the 2005 survey year certainly could be deemed a success.

3.2 2006 Survey Test Results

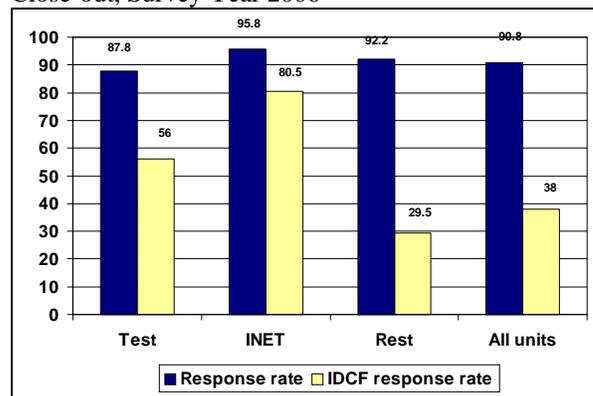
The BLS mailed out the electronic options booklet to 88,421 employers in survey year 2006 and has received a total of 93,635 responses through the IDCF and e-mail. (This number of responses include more cases than those in the test; therefore, it is higher than the mail-out number.) Of these, 36,973 were employers who received the standard booklet – 35,669 employers using the IDCF and 1,304 employers using the e-mail option. Therefore, 56,662 of the employers who received the electronic options booklet chose an electronic method to report their data – 48,800 (55.2 percent of employers receiving the electronic options) responding through the IDCF and another 7,862 (8.9 percent) using the e-mail option.

3.2.1. Performance rates

The analyses for survey year 2006 focused on the difference between those employers in the test and those not in the test (termed “Rest” in the Figures below). Since all employers in the test States were offered the electronic options, these tests were hoped to be a better illustration of how the entire nation would respond if offered the electronic options. Analyses were also performed for the INET units.

As in the test for survey year 2005, the INET units had consistently higher response rates than any other group, with a 96 percent response rate and 81 percent Internet response rate by survey close-out. The Test group lagged behind the “Rest” group in response rate: 88 percent versus 92 percent. This was also consistent with the findings from survey year 2005.

Figure 4. Response Rates by Study Group, at Survey Close-out, Survey Year 2006



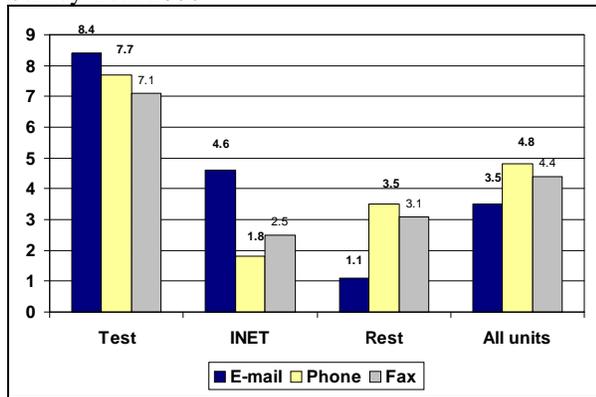
The lower response rate for the units given the electronic options versus the units receiving the standard hardcopy collection form is of concern. Reasons for this lower response rate are being evaluated and consist of:

- Since the INET (those units that have used the IDCF last year) response rates are consistently high, the lower response rates for units receiving the electronic options is a first year phenomenon only;
- The electronic options have some inherent issue that produces the lower response rate;
- The web instrument itself needs to be improved; and
- Some respondents simply prefer paper.

Analyses were also performed by industry and size class. There were no obvious trends.

In the 2006 test, the BLS was able to track other data collection methods such as e-mail, phone, and FAX. Units in the test used the e-mail option at a rate more than seven times as high than units not in the test (Figure 5). INET units also used the e-mail option more than units not in the test – the rate being more than four times as high.

Figure 5. Response Rates for Modes by Study Group, Survey Year 2006



Units in the test also used the phone and FAX much more than any other group. A hypothesis about this is that employers with questions about using the electronic options would call their State agency as directed. Once the State is on the phone with the employer, the State could take their data over the phone. Also, the State can offer the FAX form to employer’s who may wish a standard hardcopy booklet. INET units use the phone less than units not in the test and use the FAX form slightly more than units not in the test.

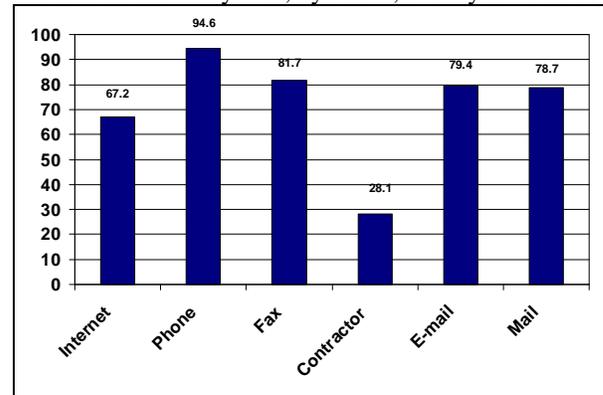
3.2.2. Processing time comparisons

With the addition of the data collection mode indicator, the BLS was able to compare the time needed for processing data between the various collection options. By tracking the time from when a sample unit’s data is received to the time that the data is edited and ready for use in estimation, any savings in processing time by using the electronic options can be determined.

With the SOII, there are many modes the respondents can use to report their data. Therefore, there are many avenues of entering their data and editing them. Data that are received via mail and keyed by State agencies are usually pre-edited (on the hardcopy) before the data is keyed into the data system. Data received by e-mail currently still need to be keyed by State personnel.⁷ The “Contractor” refers to the contractor that receives some of the mail returns. The contractor is not allowed to pre-edit data prior to keying; they just enter them into the system.

Figure 6 illustrates the percentage of returned surveys that are “clean” (ready for estimation) when the unit’s data is entered into the system. In other words, these units need no further processing to be usable. Percent “clean” is the number of units that are “clean” prior to being entered into the system divided by the total number of units received by that method. The Contractor mode has the lowest percentage clean (28 percent) because they don’t edit the data prior to keying. The highest percentage clean is with the phone mode (95 percent) because an interviewer can discuss the data with the respondent. The IDCF has 67 percent of its cases as clean.

Figure 6. Percentage of Units That are “Clean” When Entered into Data System, by Mode, Survey Year 2006



⁷ Processes to directly enter this data from the e-mail form to the database may be implemented as soon as the next survey year.

However, the editing of data takes fewer days with the IDCF (0.2 days) than with any other mode. Figure 7 shows the days needed for processing of data that are “clean” prior to data entry. The highest lag in cleaning data occurs for the surveys returned by mail to the States (13.5 days). Units in the tests took less time to reach the clean state than units not in the test. Although a higher percentage of units were clean prior to data entry, the savings in time was over six days per schedule.

Figure 7. Average Days Needed for Processing of Data That Are “Clean” Prior to Data Entry, by Mode, Survey Year 2006

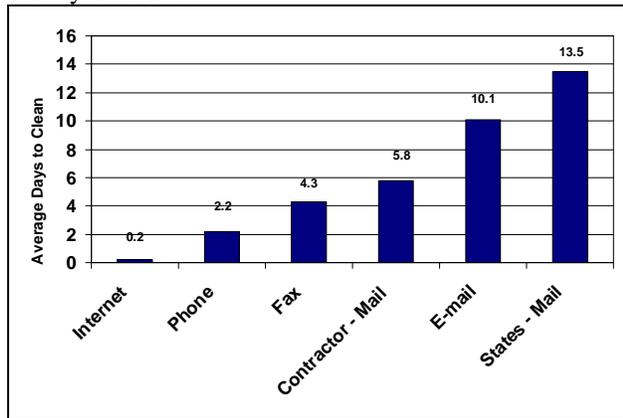
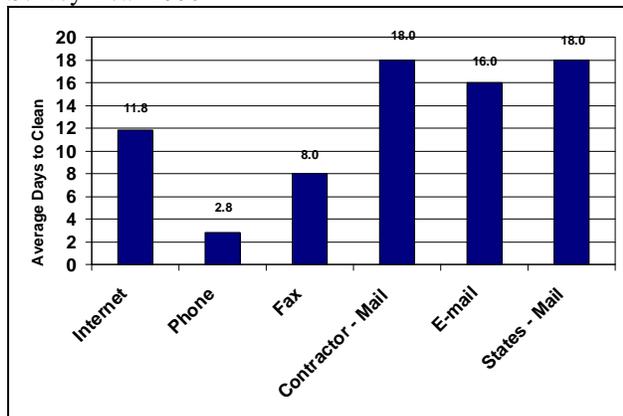


Figure 8 shows the average days needed for processing data to reach the point that it is ready for estimation (e.g., “clean”). Phone, fax, and Internet collection take the least amount of time to process. Surveys received by mail, either to a State or to the mail contractor, take the most amount of time to process: an average of 18 days for each survey.

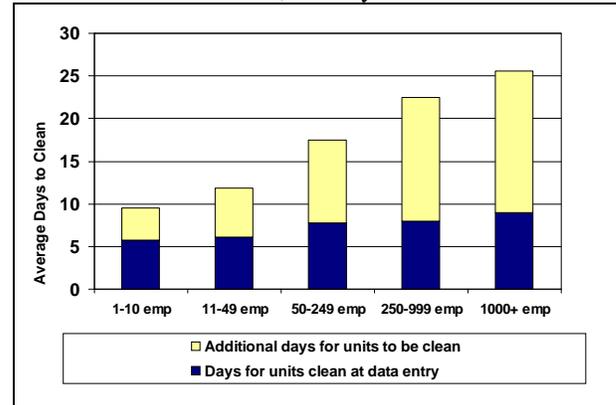
Figure 8. Average Days Needed for Processing of Data To Reach “Clean” Processing State, by Mode, Survey Year 2006



The amount of time to process units rose as the size of

the sample unit got larger (Figure 9). This was true of units that were clean at time of data entry and for all clean units. The percentages of both clean prior to data entry and total clean decreased as size class increased.

Figure 9. Average Days Needed for Processing of Data To Reach “Clean” Processing State, by Establishment Size Class, Survey Year 2006



3.3 Mailing and Printing Cost Comparisons

BLS estimated cost comparisons of the 12-page survey versus the four-page booklet (that just outlines their reporting options). These comparisons 1) assume the cost for one survey year, 2) are based on a sample of 230,000 units, 3) include nonresponse mailings and 4) assume all potential respondents receive the same data collection method. As shown in Table 1, cost savings of approximately \$257,000 could be realized by moving all respondents to e-mail or Internet.

Table 1. Potential Costs Savings of Mailing and Printing Costs

Booklet	Approximate mailing cost	Approximate printing cost	Approximate total cost
12-page standard	\$382,191.00	\$262,808.19	\$644,999.19
4-page electronic options	\$171,327.00	\$216,141.94	\$387,468.94
Difference	\$210,864.00	\$ 46,666.25	\$257,530.25

4. Discussion

Overall, the tests can be considered successes. Of those who received a booklet that described the Internet or e-mail as the primary reporting options, 55 percent reported via IDCF and another 9 percent reported via e-mail. Also, there was a reduction in processing time for the SOII data when Internet and

email are used by respondents.

Even though the survey year 2006 test can be viewed as a success in expanding the use of the electronic options for data collection, the difference between total response rates for units in the test and units not in the test is at a level that may be unacceptable. Further research is planned to identify and correct the cause of this large difference. With the savings in mailing costs, printing costs, and processing time, it is imperative that the use of electronic methods of data collection be continued and expanded.

There is a data collection test planned for SOII survey year 2007. The same 16 States that were test States in survey year 2006 agreed to be in another test. Plus, there will be two additional states.

Also, the 4-page electronic options booklet has gone through two rounds of cognitive testing plus an expert review in order to make improvements for survey year 2007. These changes:

- Highlight that SOII is required by law, as more people in this study noticed the “Your response is required by law in 30 days” on the cover page;
- Emphasize there are two reporting options – Internet and e-mail;
- Explain that e-mail reporting is an option, which wasn’t clear to respondents who had read the previous version of the instructions; and
- Make sure the instruction booklet doesn’t look like the actual hardcopy of the survey⁸.

This revised reporting options booklet will be used in the 2007 test.

Furthermore, a new e-mail option is planned. In survey year 2007, a fillable PDF form will be e-mailed to employers who request the e-mail option. (This will replace the MSWord template used for the past three years.) This form will automatically send their data to a central e-mail account for loading to the central database. States will not have to re-key the e-mailed data as in past years.

Finally, there will be approximately 27,024 INET units for 2007. This is an increase of 103.2 percent from the past survey year. These are employers who responded via the IDCF in the 2006 survey year and were selected for the 2007 survey year.

⁸ There was anecdotal evidence that respondents assumed the instruction booklet was the actual survey since they looked so similar, and therefore were less likely to read it.