March 6, 2003

Changes to Editing Strategies when Establishment Survey Data Collection Moves to the Web

By

Amy E. Anderson, U.S. Census Bureau
Stephen Cohen, Bureau of Labor Statistics
Elizabeth Murphy, U.S. Census Bureau
Elizabeth Nichols, U.S. Census Bureau
Richard Sigman, U.S. Census Bureau,
Diane K. Willimack, U.S. Census Bureau

Contact authors:
Stephen Cohen, Bureau of Labor Statistics,
cohen_steve@bls.gov, ph. (202)-691-7400

Richard Sigman, Bureau of the Census,
Richard.S.Sigman@census.gov, ph. (301)-763-4952

For presentation to the Federal Economic Statistics Advisory Committee
Bureau of Labor Statistics Conference Center, Washington, D.C.
March 21, 2003

1. Introduction

Human error, such as programming mistakes, miscalculations, keypunch errors, and interviewer misclassifications are a pervasive fact-of-life for surveys. Their contribution to the total survey error, however, may be controlled or at least somewhat mitigated by good survey practice, forethought and planning, and the advances of computer technology. All such efforts to seek out and correct these errors fall under the purview of *data editing*.

Editing may occur at almost any phase of data collection or analysis. It ranges from the almost routine activities of correcting typographical errors or out-of-range entries done by interviewers as they enter information into the computer from the field or telephone center up to the elaborate statistical checks performed by computers to identify misshapen blocks of aggregate data. In longitudinal surveys comparisons to previous results are made. But in all cases, the goal is the same: to identify and correct as much error as possible.

In mail surveys of establishments, data editing is performed during post-data-collection processing. In establishment surveys that utilize computer-assisted data collection technologies, however, some or all of the data editing can be performed during data collection. For example, in surveys that use computer-assisted telephone interviewing (CATI) or computer-assisted personal interviewing (CAPI), data editing rules, referred to as *edits*, can be incorporated into the CATI/CAPI instrument so that the interviewer is notified of a response that fails one or more edits. The interviewer can then probe for an alternative response or for verification that the response is correct.

In addition, the Bureau of Labor Statistics (BLS) uses a hybrid technique for its compensation programs—data is collected in establishments using paper and pencil techniques by trained field economists and entered after the interview on a laptop with the data uploaded to servers in the National Office. The laptop system has edits that must be addressed before the field economist can successfully call the schedule complete. Edits on the data base then flag schedules that initially looked correct but presented edit problems when viewed against other data.

Edits can also be incorporated into computerized self-administered questionnaires (CSAQs), which are delivered to the respondent by mailing disks or CD-ROMs or transmitted electronically over the Internet's World Wide Web. When a CSAQ is delivered by mail, respondents install it on their own personal computers. There are two types of web-delivered CSAQs: downloadable or browser-based. When using a downloadable CSAQ, respondents download the CSAQ from a web server to their personal computers, install it to complete the self-administered form, and then transmit data back over the Internet to the survey organization. When using a browser-based CSAQ, respondents use a web browser communicating over the Internet to view the questionnaire and to enter responses to survey questions. Browser-based CSAQs are also called *online web-survey questionnaires* or *Internet questionnaires*. When a CSAQ is used to collect survey data, respondents--not interviewers--are notified when a response fails one or more edits.

Survey organizations that collect data from establishments are interested in web-based data collection for a number of reasons. One of these is the economics. Web-based data collection

reduces mailing costs, interviewer costs, and data entry costs. Also, the use of computer-generated electronic mail in conjunction with web-based data collection reduces the costs associated with providing advance notices and reminders to respondents. Multimode surveys that include the web allow respondents to select the most suitable mode for their circumstances, in a sense, reducing burden. Web-based data collection also augments the palette of the questionnaire designer by adding to it web technologies such as hyperlinks, color manipulation, dynamic graphics, and multimedia players that can provide instructions visually and/or aurally.

The objective of this paper is to solicit advice from members of the Federal Economic Statistical Advisory Committee (FESAC) about changes in data editing strategies when survey organizations that collect data from businesses adopt web-based data collection methods. Both the BLS and the Census Bureau are in the process of incorporating web-based data collection into a number of their economic surveys and censuses. They are doing this to achieve the following potential benefits:

- Control of costs
- Improved response rates
- Decrease in perception of burden
- Improved data quality and
- For surveys with multiple closings--smaller revisions between preliminary and final estimates.

From the viewpoint of survey operations, the purpose of data editing is to provide information to an *edit reviewer*, who (or which) processes this information and decides on an appropriate follow-up activity. In interactive post-data-collection editing, the edit reviewer is a clerk or subject matter expert, whereas in fully automated post-data-collection editing the edit reviewer is another computer program. For CATI and CAPI the edit reviewer is the interviewer, and for CSAQs and online web surveys the edit reviewer is the respondent. Because the edit reviewer is usually a person, behavioral scientists may be able to provide insight into the performance of data editing with respect to the cognitive processes involved and system characteristics that limit or enhance these processes.

From the viewpoint of data quality, on the other hand, the purpose of data editing is to detect potential nonsampling errors—item nonresponse errors when data are missing and measurement errors when data are identified as inconsistent, questionable, or technically impossible. Hence, survey researchers should also be able to provide insight into the effects of changes in editing strategies in terms of the effectiveness of these changes in reducing nonsampling errors. Unfortunately, all of the data needed to study data editing from the viewpoint of data quality are not readily available in many survey-processing systems but instead must be obtained from research experiments imbedded in operational surveys. Nevertheless, this viewpoint allows us to state several survey design issues associated with changes in data editing when survey organizations adopt web-based data collection methods.

Section two of this paper describes editing practices at the Census Bureau and BLS. Sections three and four describe experiences with CSAQs and Internet data collection at the Census

Bureau and BLS, respectively. Finally, section five presents a number of discussion points for the committee.

2. Editing at Practices at the Census Bureau and BLS

2.1. Editing at the Census Bureau

The Census Bureau has developed two generalized processing systems for editing economic data—one for the economic censuses conducted every five years and a second for more frequently conducted current economic surveys. Both of these systems can be configured to process a particular census or survey, without the need to rewrite the system's computer code. The developers of the economic census editing system called it "Plain Vanilla" because it provides basic editing capabilities, which can be augmented, if necessary, by trade-area-specific computer code. The current-surveys editing system is a module of the Standard Economic Processing System (StEPS), which performs a number of post-data-collection processing functions. The development of Plain Vanilla and the StEPS editing modules were similar--user requirements arose from experiences with earlier systems, and the primary objective for both development efforts was replacing multiple editing systems with a single system. These development activities differed, however, in that they developed different types of systems: a highly automated system for the large data volumes associated with economic censuses and a very flexible and easily configured system for the large number and large variety of current economic surveys.

Currently, the Census Bureau uses StEPS to process approximately 80 current economic surveys. StEPS has a module for data editing and two modules for imputation. The two imputation modules perform what StEPS labels simple imputation and general imputation. The usual order in which the modules are executed is first simple imputation, then editing, and finally general imputation. The StEPS simple imputation module imputes data values considered to be equivalent to reported data. The resulting data are flagged as being reported. A frequently performed type of simple imputation is "data filling;" i.e. StEPS fills in missing data that can be easily inferred from other data. Another type of simple imputation is adjustment of two or more details items that add almost but not exactly to a corresponding total item. The StEPS editing module performs automated detection of potential data errors, which are data that individually or in relationship to other data fail to conform to expected reporting behavior. The editing module allows subject-matter analysts to interactively define edits and then to execute defined edits interactively or in batch. Analysts can define the following types of edits:

- Required-item edit. Verifies that a specified item has been reported.
- Range edit. Verifies item value lies in the range defined by a specified minimum and maximum value.
- List-directed test. Verifies the value of the specified item is contained in a predefined list of values.
- Balance test. Verifies that a sum of specified detail items is equal to a specified total.
- Survey-rule test. Free-form test that validates complex interitem relationships.
- Negative test. Verifies that the value of the specified item is not negative.

The StEPS editing module only identifies the edit failures; it does not change data. Executing edits interactively allows analysts to interactively make data corrections and then see if the corrected data satisfies the edits. Executing edits in batch creates an edit reject file, which can be reviewed interactively or passed to the general imputation module to identify data to be machine imputed. For more information about the StEPS editing module, see Sigman (2001).

The development of Plain Vanilla prior to the 1997 economic censuses allowed the Census Bureau to replace four separate editing systems with a single system. Plain Vanilla has modules for performing ratio, range, and balancing edits and a generalized module for validating discrete data items against item-specific reference lists. To support the Plain Vanilla ratio edit module, the Census Bureau has developed software to generate implicit edits in accordance with the Fellegi and Holt (1976) editing approach and software that determines upper and lower limits for the ratio edits by calculating "resistant fences" (based on medians and quartiles) from prior-period data. The Plain Vanilla ratio-edit module analyzes the pattern of ratio edit failures, determines the weighted minimum number of items that have to be changed to satisfy all ratio edits, and then imputes these items using a hierarchy of models such that the imputed values satisfy all the specified ratio edits. Similarly, the other Plain Vanilla modules examine patterns of associated edit failures and then in accordance with specified models impute data values that satisfy the specified edits. Subject-matter experts monitor the editing process and, if necessary, make adjustments to edit parameters or override edit actions. For more information about Plain Vanilla, see Wagner (2000).

2.2. Editing at BLS

BLS processes each of its surveys on systems that are individually tailored to the survey design to ensure maximum efficiencies in the agency's operations. All BLS survey data are extensively edited to ensure the agency releases the most accurate statistics possible.

In addition to verifying valid entries, each post processing systems includes some of the following type of edits:

- Linear edits (i.e., their graphs are straight lines)
- Non-linear edits
- Conditional edits
- Ratio edits
- Variables that can accept negative values (negative values are edited out)
- Quantitative edits
- Qualitative edits
- Univariate range checks
- Multivariate, record-level checks for consistency
- Other consistency checks (e.g., work schedule if full-time employee)
- Route checks or skip patterns and
- Check of generated values against external standards.

See Appendix A for a complete summary of edit practices at BLS.

As advances have occurred in technology, data collection has progressed away from collecting information by paper and pencil, then keypunching the data onto a database. Current technology permits the field representative to enter data directly on a computer and then transmit the data to a database. EDI allows us to copy respondent data files directly to a database. For surveys with only a few data elements, respondents can enter data directly into agency databases through touchtone data entry. Mail surveys are processed by keypunching the data, editing for keying mistakes and then editing for logical errors.

3. Census Bureau experiences with CSAQ data collection

The Census Bureau consists of several directorates that conduct censuses and surveys. The Economic Programs Directorate conducts economic censuses every five years and conducts current economic surveys monthly, quarterly, and annually in areas of manufacturing, construction, commercial services, government services, and foreign trade. The Methodology and Standards Directorate conducts research that supports the survey activities of the other Census Bureau directorates. In this section we describe the experiences of the Economic Programs Directorate in incorporating edits into CSAQs, web-CSAQs, and online web-survey questionnaires in the 2002 Economic Census and in several current surveys, and we describe supporting research in this area conducted by the Methodology and Standards Directorate.

3.1. Survey of Industrial Research and Development

The Survey of Industrial Research and Development (R&D Survey) is sponsored by the National Science Foundation and conducted annually by the U.S. Census Bureau. It is the only complete source of national data on R&D spending by industry and on the number of scientists and engineers in industry R&D. The target population consists of all industrial companies that perform R&D in the United States, and the current sample size is approximately 32,000 companies. All items on the self-administered survey form (see http://www.census.gov/mcd/resdev.html) are mandatory for the 2002 survey year. (The Office of Management and Budget approved mandatory authority for all items in 2002 in conjunction with the 2002 Economic Census.) Results are used by industry, government, and academia to evaluate the state of science and technology in the U.S. and to develop government and corporate policies.

The R&D Survey uses two different questionnaires: Form RD-1 is sent to companies that reported $3 million or more in R&D expenditures in the previous year's survey, and Form RD-1A is sent to the remaining companies in the R&D sample. Form RD-1A contains only five major items, whereas Form RD-1 contains 13 major items. The major items in Form RD-1 have many parts, however, so that there is a total of 190 data fields associated with Form RD-1. The mode of data collection for Form RD-1A is mail only, while Form RD-1 sample cases are currently offered both mail and CSAQ response modes.

Sweet and Ramos (1995) analyzed edit and respondent burden data from the 1994 R&D diskette-by-mail CSAQ pilot study. In this study 200 companies, which were similar in size, hardware/software capability, and had indicated interest in trying a CSAQ, were assigned to either the paper or the CSAQ panel. Significantly fewer edit failures were found during the post-data-collection processing with the CSAQ respondent data than with the control panel (paper mode) data. Twenty-nine item data edits consisting of balance, interitem, missing, and logical were run in the CSAQ. The addition of edits did not appear to increase respondent burden unreasonably. The inclusion of edits was not mentioned as a reason for nonresponse as found during a telephone follow-up of CSAQ panel nonrespondents, although four of the 32 CSAQ nonrespondents polled mentioned that using the computer would make it more difficult. This was not elaborated.

In this experiment, two respondent burden measures were evaluated: voluntary satisfaction study responses and time-on-task estimates. Most respondents were satisfied or very satisfied with the CSAQ reporting system. "Resolving errors" and "making comments" were areas that were noted as neither particularly easy nor difficult. No one mentioned resolving errors or edits in the open ended remarks, although changing data was difficult since navigation within the screen was difficult (this was not a Windows application).

Although both panels were asked to keep track of their total time to complete the survey including preparation/gathering of data and reading instructions, only two respondents in the control panel provided that information. In the CSAQ panel, respondents estimated that completing the electronic questionnaire took about 18 percent of the total time needed, on average. In addition the CSAQ instrument logged the amount of time it was "operating." The average was 86 minutes, which is in keeping with the respondents' self assessment and well within the OMB burden hour estimate of 20 hours.

Additional item-data edits have been incorporated into the R&D CSAQ since 1994. Currently, it contains 57 item-data edits, broken down as follows:

- 2 interitem ratio tests (ratio of sales to employment, and ratio of total R&D to sales)
- 23 current-to-prior ratio edits
- 36 balance edits and
- 3 zero-item edits, which ask the respondent if an entered zero is correct.

During post-data collection processing, the Form RD-1 data from both CSAQs and mail returns are edited to identify missing or questionable data. There are 260 post-data collection edits, broken down as follows:

- 191 missing-data edits
- 8 logical edits
- 4 magnitude edits
- 7 interitem ratio tests (3 of these edits involve the two interitem ratios tested in the CSAQ)
- 15 current-to-prior ratio edit, (all 16 of these are contained in the CSAQ) and
- 36 balance edits (all 36 of these are contained in the CSAQ)

Two of the inter-item ratio edits performed during post-data collection processing are also performed in the CSAQ. The CSAQ's error messages for these are: the following:

- "The ratio of Sales to Employment is unusual" and
- "The ratio of Total R&D to Sales is unusual"

On the other hand, there are five interitem ratio edits performed in post-data collection processing that are not contained in the CSAQ. These are for the following ratios:

- (# R&D scientist and engineers) / (# employees) {edit failure if ratio > 1.0}
- (# employees) / (total R&D)
- (# R&D scientists and engineers) / (total R&D)
- (company R&D for current year) / (budgeted R&D for next year)

In the first three of these edits, each item appears in two different edits. These three edits constitute what Felligi and Holt (1976) refer to as a *complete set of edits*, having the mathematical property that one can analyze the pattern of resulting edit failures to determine which items must be changed in order for all the edits to be satisfied. This analysis process is referred to as *error localization*. The Census Bureau uses the Felligi-Holt approach to develop ratio edits for its Economic Census, but does not explicitly use this approach for developing ratio edits for its current economic surveys. The first three interitem ratios in the above listed were specified by R&D analysts when developing their post-data collection edits. Presumably, with practice they learn how to interpret the pattern of resulting edit failures to determine which item (or items) is most likely to be questionable—that is, they learn how to mentally perform error localization. However, such an analysis would likely be cognitively challenging for CSAQ respondents to perform, as well as difficult to communicate via edit messages. Nevertheless, error localization can be programmed, which raises the question of whether CSAQs should contain complete sets of ratio edits, plus the ability to perform error localization.

3.2. Manufacturer's, Shipments, Inventories, and Orders Survey (M3)

Each month, the Census Bureau conducts the Manufacturer's Shipments, Inventories, and Orders (M3) Survey, which obtains measures of current industrial activity and indicators of future production from U.S. manufacturers. The M3 survey is a voluntary nonprobability panel survey for which nearly all U.S. manufacturing companies with $500 million or more in annual shipments provide data. Smaller companies also report data to improve coverage in selected industry groups. The industry categories fall into one of two major groups: durable goods, such as motor vehicles and aircraft; and nondurable goods, such as apparel and food products. The sample includes a total of about 3500 company reporting units, where a reporting unit consists of all company operations in one industry category. Thus, large diversified companies often have multiple reporting units.

In contrast to the R&D survey described in section 3.1, M3 is a very short, straightforward survey consisting of five or seven survey items, depending on the industry category of the reporting unit (see http://www.census.gov/indicator/www/m3/contactinfo/instructioninfo.htm for a copy of the form). For most nondurable goods categories there are five items: value of shipments, total inventories, and stage of fabrication (materials and supplies, work-in-progress, and finished goods). M3 also requests new and unfilled orders from reporting units in the durable goods

categories and in a few nondurable goods categories. All respondents are also asked to indicate whether their reported data are for a calendar month, a four-week period, or a five-week period.

Respondents may report via mail, fax, touch-tone data entry, or voice recognition. Fax returns are digitally processed using character-recognition algorithms to extract reported data. Survey staff enter mail returns into a fax machine and process them as if they had been faxed by the respondent. Consequently, there is no need for manual keying of mail-return data by Census Bureau staff; however, staff need to visually inspect each of the forms to ensure that the character-recognition software has completely captured the data.

In 2000, the M3 survey began web-CSAQ data collection from 100 reporters. This was preceded by usability testing with six internal Census Bureau staff members. This early design used server-side edits, where a separate page displayed the edit failures once the CSAQ was submitted. Users had to exit that page to return to the form and correct their responses. Users had two complaints. They wanted to see their errors, like a summation error, immediately. They also wanted to be able to navigate back and forth between the errors and the form. Usability experts recommended immediate edit messages over server-side edits; but if server-side edits were the only option, usability experts recommended placing the edit messages directly on the form, eliminating the need to navigate between two windows.

In July, 2002, a revised version of the web M3 was rolled out to the original 100 Web reporters. Along with new security procedures, enhancements included the following features:

- Additional basic edits have been built into the web-CSAQ
- Respondents can choose to mark their form "unfinished (with errors)"
- Additional edit that checked the consistency of the reporting period

Server-side edits remain in the new version of M3's electronic form. After completing the form, respondent must click a button to save their information. If the data contain errors, the survey form re-appears containing edit failures. The respondent is notified by a colored box located at the top of the screen, which contains an error icon and a message stating that errors were found in their data. The respondent must then scroll through their form until they reach the field or fields that contain colored messages. These messages contain information about the cause of the error. Once respondents have addressed each of the problems, they must save their data again. If errors persist, respondents must continue to address them or choose to move on, leaving errors intact. Respondents are allowed to submit their forms with errors.

The current M3 CSAQ includes the following item-data edits:

- Reporting-period information (i.e., reporting for a month, four weeks, five weeks, etc.) not consistent with respondent's prior reporting practices
- Data value negative for an item that does not allow negative data
- Large disagreement between reported new orders and derived new orders
- Large disagreement between total inventories and the sum of stage-of-fabrications inventories
- Reported value less that the minimum of previously reported data for a particular month or greater than the maximum of the previously reported data

- Alphanumeric characters in a field that do not allow for that type of character.

Starting in late spring of 2003, the Census Bureau will use its Standard Economic Processing System (StEPS) to process M3 data. For the M3 survey, analysts defined seven edits for purposes of identifying missing data to be machine imputed and defined 40 item-data edits for purposes of identifying questionable data. When a data item fails one or more of these latter edits, it is flagged for exclusion from estimation. Analysts then review edit-failing items and can decide to reset the status of an item to be included in the link-relative calculation of the prior-month-to-current-month percentage change for the industry category or to be treated as an outlier. (Outliers are excluded from the calculation of the link relative but by representing only themselves are included in the calculation of the current month's estimated universe level.) The M3 post-data-collection edits for purposes of identifying questionable data include those incorporated in the CSAQ, plus the following additional edits. (If the stated condition is true, then the associated data item is flagged as failing the edit.)

- Item's current-month data equals its prior-month data
- Reported value less than a preset variance from the minimum of previously reported "good" data for a particular month or greater than a preset variance from the maximum of the previously reported good data for a respondent that has reported at least three years of good data for a given month over the last five years
- Item's average over the last three years of prior-month-to-current month percentage change or of absolute change for the given month outside a plausible range for such three-year averages for the respondent's industry
- Item's prior-month-to-current-month percentage change for the given month outside a plausible interval about the industry-level average percentage change for that month predicted by time-series seasonality modeling
- Reported data for an item that has not had reported data in the past two months.

Generally, these additional post-data-collection edits--i.e., those not incorporated into the M3 CSAQ--involve comparisons of reported data to various summary statistics calculated from historical M3 data.

3.3. Summary of CSAQ edits and implementation strategies

Edits for CSAQs are devised from the perspective of respondents, taking advantage of the opportunity to ask respondents to clarify or resolve problematic information. Any fields containing information that respondents can touch may be subject to an edit check. These fields include name, address and contact information along with data items. Thus the types of edits incorporated into CSAQs may cover a broader range of potential errors than those conducted during post-collection. The types of CSAQ edits incorporated into Census Bureau CSAQs include:

- Balance edit. Verifies that the sum of detail data equals the appropriate total
- Preventative edit. Blocks respondents from completing an action, occurring upon the first invalid keystroke

- Ratio edit. Verifies that the ratio of two data values lies in the range defined by a specified minimum and maximum value lies within a specified ratio
- Logical edit. Verifies that data reported for related items are logically valid
- Alphanumeric edit. Verifies that the data meets the proper alphanumeric rules established for that field
- Missing value/Incomplete edit. Verifies that data has been reported
- Format edit. Verifies that the data has been entered in the expected form (i.e., date entered with dashes instead of slashes)
- Rounding Test. Checks to see if rounding has occurred in the data.

Survey designers also have control over when and how the various edits appear to respondents. Immediate edits respond instantly upon detection of erroneous data, and the system prompts the user for a correction or explanation. Immediate edits are presented as either a pop-up window or an icon located near the questionable item. Deferred edits are presented to the respondent after the data has been entered and reviewed, usually in a list format.

Table 1 summarizes the editing features of the R&D survey and the M3, along with four other Census Bureau programs offering downloadable or browser-based CSAQs – the Company Organization Survey (COS), the Annual Survey of Manufactures (ASM), the Quarterly Financial Reports (QFR) and the 2002 Economic Census.

Table 1. Summary of CSAQ edits

	Type of CSAQ	**Ratio of edits to fields**	**Types of edits***	**Timing of the edit messages**	**How displayed**	**Resolution required to submit?**
R&D	Downloadable	67/205=.33	P, R, M	Immediate, Deferred	Review panel	N
M3	Browser -based	103/58=2.43	P, B, R, A, M	Deferred	Highlighted text	N
COS	Downloadable	36/23=1.57	P, R, L, M, F, RT	Immediate, Deferred	Pop-up messages, review panel	Y (for a few key edits)
ASM	Downloadable	unav/88	unav	Immediate, Deferred	Pop-up messages, review panel	N
QFR	Downloadable	29/94=.31	B, P, L, M	Immediate	Icon next to item	N
2002 Economic Census	Downloadable	66/95=.69	B, P, L, M, F	Immediate, Deferred	Icon next to item, review panel	Y (for one edit)

*B=Balance, P=Preventative, R=Ratio, L=Logical, A=Alphanumeric, M=Missing value/Incomplete, F=Format, RT=Rounding Test

The following describes key features of the editing strategies incorporated into these electronic data collection instruments, along with the rationale behind them.

Research and Development Survey (R&D): The R&D CSAQ edits were taken directly from the survey's post-data collection editing system. The original CSAQ included only balance edits. Over the years, as the CSAQ evolved, additional edits were incorporated. At first, this included edits for problems typically addressed in analysts' callbacks to respondents. Eventually, the edit checks grew to encompass nearly all of the post-data-collection edit rules.

R&D respondents are able to submit their data with unresolved edit failures. Because of longstanding relationships with many respondents, R&D staff did not want to discourage companies from responding. The R&D staff was concerned that respondents may become irritated at being stopped by edit checks, and that this could affect decisions to respond to the survey at all.

Manufacturers Shipments, Inventories, and Orders Survey (M3): Edits originally programmed into the 2000 M3 CSAQ form were enhanced for the current web version. Analysts chose the web CSAQ edits based on what seemed necessary and logical. Numerical parameters selected for the web CSAQ ratio edits, which involve comparisons between current and prior period data, were chosen arbitrarily. These current-to-prior ratio tests result in an edit failure if the current figure is less than one half of or greater than twice the prior figure. The same limits are used for the current-to-prior ratio test in post-data-collection processing. A unique edit built into the M3 permits minor rounding errors that often occur for the question requesting "new orders." In addition, since M3's server side edits cannot prevent a respondent from entering invalid characters, alphanumeric edits are included to alert respondents to unacceptable responses.

None of M3's CSAQ edits will block respondents from submitting their data. When making this decision, analysts looked at the trade-offs between "getting the data and getting good data." They decided to obtain data from respondents in any form possible and to handle problems after receipt.

Company Organization Survey (COS) and the Annual Survey of Manufactures (ASM): The Company Organization Survey is an annual survey that obtains current organizational and operating information on multi-establishment companies in order to maintain the Census Bureau's Business Register. Companies are asked to report first quarter payroll, annual payroll, and employment for each establishment along with identifying establishments that have been sold, closed, started, or acquired. The Annual Survey of Manufactures provides detailed annual statistics on the location, activities and products of U.S. manufacturers. The survey obtains basic data such as kind of business, ownership, cost of materials, inventories, value of shipments, payroll and employment.

CSAQs for both surveys were originally developed and are managed by the same Census Bureau staff. Over the years, the edits in the COS and ASM have remained fairly static and consistent. Survey staff annually request minor updates to the instruments, including additional edits.

For the ASM and COS, the minimum level of information required to consider the form complete was a factor in determining which edits were included in their CSAQs. The result was to assign edits to the most critical items, e.g., employment, payroll and operational status; these are the

same edits considered most critical in the paper environment. In the COS, numerical editing parameters were estimated from historical data using statistical methods and algorithms were created based on historical knowledge.

In the COS, there are just a few so-called "hard" edits for which unresolved failures prevent data submission. These were assigned to items required to be consistent with one another within the form, and these edit messages are displayed after the last of the associated items is entered. The number of hard edits grew over time as staff experience and respondent acceptance grew, which was enhanced by longstanding staff relationships with COS reporters. None of the ASM edits must be resolved before data submission.

Quarterly Financial Report (QFR): The Quarterly Financial Report is a quarterly survey that obtains current statistics on domestic corporate financial conditions. Companies provide standard income statement and balance sheet data including sales, depreciation, cash, accounts payable, liabilities, and stockholders equity.

The QFR survey assigns three levels of severity to their CSAQ edits: information, warning and error. The level of severity was determined by the importance of that item, with mission-critical items, such as assets, liabilities, and the satisfaction of certain accounting tautologies, being assigned "error" status. Range edits were not programmed into the QFR CSAQ because the same instrument is sent to companies of all sizes and industries. Building "one-size-fits-all" parameters for this survey would result in meaningless edits for many industries.

The QFR CSAQ does not prevent respondents from submitting data with unresolved edit failures. The QFR staff is concerned that hard edits would discourage companies from responding to the survey. The staff would prefer dealing with data inconsistencies in post-data collection than risk losing a respondent. In future CSAQ versions, they hope to incorporate some hard edits for one item, which, if missing or incorrect, causes the entire form to be unusable.

Economic Census: The Economic Census provides the foundation for most of the United States' economic statistics, and the benchmark for the Gross Domestic Product (GDP) and other leading indicators of economic performance. The Economic Census is collected every 5 years, for years ending in "2" and "7" and includes every industry and geographic area.

For the first time, all U.S. businesses in the economic census have been offered the opportunity to report electronically via CSAQ. When developing the edits for the 2002 Economic Census, staff first reviewed the edits included in the 1997 Retail Census CSAQ, which were applied and expanded for 2002. The Style Guide for the 2002 Economic Census Electronic Forms (Economic Electronic Style Guide Team, 2001) was also consulted when selecting the types of edits and the wording of associated edit messages.

The original specifications provided by the various trade areas included numerous edits. Due to concern for respondent burden, staff was asked to choose only the most important edits. Staff then identified the most critical items, such as employment, payroll, sales/revenues, and created edits for those items.

Software limitations and schedule constraints also affected the types of edits selected for the economic census CSAQ. Since many edits were built to cover all 550 economic census forms encompassing a wide variety of industries, they were kept at a very general level. Thus, range edits were not included in the economic census forms, because, like QFR, "one-size-fits-all" parameters were considered uninformative.

Only one edit in the 2002 Economic Census CSAQ prevents respondents from submitting establishment data if it is not resolved. This item, the number of months the establishment operated during 2002, was deemed the most critical relative to the objectives of the economic census. For the remaining items, respondents can submit their data with errors. The staff kept hard edits to a minimum because they prefer to receive "bad data rather than no data." This philosophy appears prevalent at the Census Bureau.

3.4. Findings from Usability Research

A wide range of edit behaviors have been tested and used in Census Bureau CSAQs. "Behavior" here refers to the CSAQ's methods of communicating edit failures to the respondent: What is communicated, how is it communicated, and when? To learn about respondent interaction with various edit behaviors, the Census Bureau tests candidate editing approaches with respondents. To assess the edit behavior design, we observe how the respondent interacts with the edit design during usability testing. Do respondents recognize the edit failure notification? Do respondents read the edit messages? If they read the messages, do they understand them? What type of action do they take regarding the message? A response might consist of ignoring the edit, modifying data values, or writing a remark to explain why data are accurate even though they failed the edit. The latter task is particularly characteristic of business surveys since valid data often lie outside an expected range. Finally, how easy is it for respondents to interact with the CSAQ to respond to the edit? For example, they may have to navigate to the item referred to in the edit.

We have found it virtually impossible to recruit business respondents to travel to the Census Bureau for usability testing in the Census Bureau's state-of-the-art usability lab. Instead Census Bureau staff travel with a video camera and laptop to the businesses to conduct usability testing. Video taping (with the respondent's consent) allows one or more researchers to analyze the session afterwards; the laptop is a necessary backup in case the CSAQ does not work on the respondent's workstation. Using the think-aloud protocol often used with cognitive testing, we watch and listen as respondents use the CSAQ in their offices.

Usability testing has its limits, however. Since usability testing uses a small number of subjects, results cannot be tested for statistical significance. Usability testing is not an experiment. It is intended to identify problems that actual respondents might have with the CSAQ, not to find significant differences between groups. Compounding this limitation is the difficulty in recruiting business respondents for this type of testing. Thus, even though all of the following economic surveys (COS, ASM, QFR and M3) have had some amount of usability testing, very often there has only been a handful of actual business respondents in each study. Often we use internal staff

members as supplemental subjects, since usability testing typically recommends between 6 to 12 participants (Dumas and Redish, 1999).

Also usability testing does not focus solely on edits. Instead it focuses on the entire instrument, with a limited amount of time devoted to edits. Edit functionality is typically one of the last features programmed. Thus, edit behavior is often not fully functional when usability testing is conducted, as seen during usability testing of the 2002 Economic Census prototype. With those disclaimers in place, we urge the reader to realize that the best practices identified for edit behavior arise from limited usability testing and need to undergo additional research.

Several themes regarding editing emerged during our review of the internal business survey usability reports. The usability reports for two institutional surveys Web-CSAQs, the Library Media Center Survey Field Test (LMC) and the Private School Survey (PSS), confirm some of those themes and are included in the analysis below. Although two rounds of testing occurred in preparation for the creation of the 2002 economic census electronic form style guide (Economic Electronic Style Guide Team, 2001), only three local companies tested the prototype RT-44401 (Building Materials and Supply Dealers) CSAQ for usability in the summer of 2001. Unfortunately no edits were embedded because the survey development software was not functional at the time of the test. We have excluded results from that report (Nichols et al., 2001b). For reference purposes we share the number of participants in each of the tests and when the test was conducted:

Table 2. CSAQ usability testing

Data Collection Instrument	Number of Respondents	Date of Testing
COS	3 current COS respondents	September 1998
M3	6 Census staff	December 1999
LMC	13 school librarians	June-July 1998
ASM	3 current ASM respondents 3 Census staff	February 2000
QFR	3 current QFR respondents	July 2000
2002 econ census edit wording test	10 Census staff	Spring 2001
PSS	8 private school employees	2001

The following design guidelines summarize our interpretations of the results from these studies of edit behavior

1. Good questionnaire design minimizes edit failures. Good questionnaire design includes communicating to respondents what data fields are required. For example, instructions should inform participants to click on “none” if the answer is zero or to enter a number when an entry is required (Bzostek and Mingay, 2001). Formatting errors should not arise unless respondents ignore instructions or miss the required format. For dates or amounts, format can be built into fields automatically. Additionally, question text can include

instructions on the correct format (Bzostek and Mingay, 2001, Nichols et al., 2001a).

2. Perform edits immediately unless checking for missing data or performing an interitem edit. Defer activating those edits. Run them either immediately before the respondent finishes the form or after all the items in the interitem edit have been entered. Participants' preference is for immediate notification of edit failures, rather than receiving a list of edit failures at the end (Bzostek and Mingay, 2001, Rao and Hoffman, 1999). Participants can learn to avoid similar mistakes if they are notified immediately. However, we caution against "jumping the gun." In two different survey tests, users thought edits were ill-placed. For example, one QFR edit that was checking the consistency between two items, flagged the error as soon as data were entered for the first of the two items. This edit should have been invoked on the second of the two items. More ideally, we recommend activating the edit when all the relevant fields have been manipulated, no matter the order of manipulation (Nichols et al., 2000). Authors of business survey edits have to be very careful not to make too many or too few assumptions. For example, some participants did not think the edit took in all the relevant factors when calculating ranges (Saner et al., 2000).

3. Mixing editing with other functions, such as submitting, violates user expectations. Problems arose in the PSS because the server-side editing process was invoked when the respondent pressed the "Finished" button. All participants believed that when they pressed the "Finished" button the form would be submitted. When they saw that an error check was run and error messages appeared, they changed their understanding of the "Finished" button. They then believed that clicking on "Finished" again would iteratively check for errors until their form was correct. This was not the case. Edits were only run the first time the "Finished" button was pressed. This design, most likely created to ensure respondents invoked the edits, violated respondents' understanding twice. Initially it violated their understanding of the word "Finished." It then violated their expectation of the ability to iteratively check for errors (Bzostek and Mingay, 2001). During usability testing of the COS, respondents were also surprised that the error report was rerun when they tried to submit (Nichols, 1998).

4. Allow edit failure reports to be run iteratively. The batch process of presenting edit messages (either surver side or review screen) is not a problem in itself. The problem arises if the CSAQ does not allow this batch processing to be rerun and updated. For example, the PSS was designed for all the error messages to appear together, at the top of the form, once the form was submitted. Most likely designers thought respondents would make their way through the list, correcting each one in turn. During usability testing, however, some participants wanted to recheck their form after correcting only one error, hoping that the list would reappear without the error they had just corrected. In the ASM, we also observed respondents wanting to return to the review screen after correcting a failure. Luckily, each time the review screen was invoked, the edits were rerun, generating an updated list (Saner et al., 2000).

5. Allow for easy navigation between an error report and the items. In both the COS and

ASM, respondents easily navigated from the error report to an item by clicking on the hyperlink edit failure text. Once at an item, however, returning to the review screen was confusing (Saner et al., 2000). PSS users wanted to be able to easily navigate back to the list of edit failures, once they were at an item. When they discovered the list was at the top of the form, they complained about having to scroll up to see the list (Bzostek and Mingay, 2001). In the 2000 M3 CSAQ design, server-side edits were run and appeared on a separate page. Users had to exit that page to return to the form and correct their responses. They could not easily navigate between the two pages. Usability experts recommended placing the messages directly on the form, eliminating the need to navigate between two windows (Rao and Hoffman, 1999).

6. Clearly identify which edit failure goes with which item. For example, in the PSS, clicking on the edit-failure message reset the page to display the data entry field which failed the edit. The page was reset so that the line containing the data entry field was at the top. The question text for this field was then above the fold. To see the question text, respondents had to scroll up. In the LMC, the immediate pop-up edits were invoked when the respondent's cursor gained focus in another data field. If respondents scrolled down the form, the item with the failed edit could be off the screen when the pop-up message appeared (Nichols et al., 1998).

7. Edit messages should contain a location, a description of the problem, and an action to take. Respondents were always trying to decipher where the error was located and what they needed to do to fix it. Every participant for the 2002 economic census prototype testing commented that many of the edits would have been easier to work with had the item number been available. Participants also wanted to know what action they needed to take to resolve an edit failure. The easiest messages to understand were those that said, "Please complete item x" (Nichols et al., 2001a).

8. Edit messages should avoid jargon, be polite, use good grammar, be brief, use active voice, use words that match the terminology used in the question, and avoid offering only one of many possible solutions. Problems arose when words used in the edit message did not match the terminology used in the question. Respondents were not sure they were on the right question. Problems also arose when solutions such as, "Check for a typing mistake" were in an edit message. Sometimes these solutions led respondents awry (Nichols et al., 2001a).

9. Inform respondents about the submitting/edit failure policy. Respondents in both the ASM and QFR testing were not sure whether they could send data with errors remaining, although this was permissible (Saner et al., 2000, Nichols et al., 2000).

10. Give the respondent as much control as possible. The user is not in control with pop-up edit failure messages. Unsolicited pop-up windows containing edit failures were a problem for respondents in usability testing for the LMC Field Test. A number of respondents did not read the message but automatically clicked a button in the window to make it disappear. When probed at the end of a testing session, one respondent didn't even remember any pop-up windows. Others thought it was a computer bug. (Nichols et

al., 1998; Tedesco et al., 1999).

11. Use icons with caution since many do not have universal meanings. Both the QFR and the ASM use icons to immediately notify the respondent of an edit failure. The red circle with the white "X" icon was used successfully by the QFR respondents. They rarely occurred during the ASM testing. The yellow "!" warning messages, however, were rarely clicked on in the QFR testing, and a few ASM respondents were unaware of their clickable property. The white bubble "i" icon was only used in the QFR. When probed, respondents thought the "i" bubble icon meant additional information and were surprised to find the message reminded them to make a remark (Nichols et al., 2000).

3.5. Summary of Census Bureau Experiences — What We've Learned

Several themes seem to emerge from the Census Bureau's experience and research on incorporating editing into CSAQs and web instruments. First, early CSAQs incorporated only a few basic edits because of grave concern for additional respondent burden, which might result in total nonresponse. In addition, the early software could support only a few simple edits. Over time, more edits have been added to existing CSAQs and to newly created CSAQs. Indeed, the ratios of the number of edits to the number of questionnaire items presented in Table 1 seem high – the most recent Web instrument developed by the Census Bureau, the M3, has more than two edits for every item on the questionnaire. This growth is, in part, because of enhancements to the software, enabling the creation of edits that were not previously possible. Moreover, the number of edits has increased as survey staff experience and confidence have grown over multiple survey iterations.

Indeed, even though the number of edits has increased, it appears that embedded edits do not necessarily lead to unit nonresponse. This is corroborated by usability research that suggests that respondents seem to appreciate the edits, wanting to be informed of discrepancies in their data so they can make corrections. Thus, respondents do not necessarily consider edits to be "bad" and they do not appear to abandon the response task just because they received edit messages.

The source of edits added to CSAQs are those typically applied during post-collection processing. It appears, however, that some, but not all post-collection edits may be moved into the CSAQ environment. The reasons for this vary. One limitation is the software--for all their enhancements, there remain programming or technical issues that constrain development of embedded edits. This was one of the reasons fewer edits were incorporated into the 2002 Economic Census, because the design of the system inhibited some edits being built. The utility of some edits is also limited by "one-size-fits-all" approaches to the design of the electronic instrument. That is, the correctness of many establishment survey data items depends on the industry, so editing parameters vary by industry. CSAQs are not currently tailored by industry or size of the business, limiting the value of certain types of edits.

Because of various constraints, survey managers at the Census Bureau must prioritize edits incorporated into CSAQs. Priorities are placed on items deemed mission-critical. Subject area

knowledge of respondents' abilities to report particular data items and typical levels of response accuracy also guide defining and selecting edits for CSAQs.

Nevertheless, respondent reaction remains a valid concern. Research shows that, to a great degree, instrument control needs to remain with respondents. Usability research suggests a number of guidelines for user-centered design and implementation of CSAQ edits so that the perception of added burden is minimized. Operational experience suggests that respondents easily accept edits that ensure that the data they enter meet required formats, and these types of edits are effective. In addition, different levels of edits – information, warning, error – provide respondents with information about their severity, and let respondents choose how to deal with them.

Usability research suggests that the issue of respondent control over resolving edit failures is perhaps most critical at the data submission stage. All current Census Bureau instruments allow respondents to submit completed electronic survey forms with data that have failed the embedded edits. The main reason for this strategy, however, is so not to discourage survey response due to unresolved edit failures. All survey programs prefer edit-failing data to no data (unit nonresponse), and continue to rely on post-collection editing and imputation to cleanse reported data. Thus it appears that survey managers are choosing measurement error in the collected data over nonresponse error.

This raises the question of whether "submission with errors" is a satisfactory, cost-effective, "optimum" strategy in terms of data quality, which is affected by both nonresponse error and measurement error. Can we ever realize the data quality benefits of embedded CSAQ edits if we permit respondents to submit data with errors? In fact, the few investigations into data quality conducted so far suggests that benefits of CSAQ edits are indeed realized, resulting in fewer data items failing post-collection edits (Sweet and Ramos, 1995) and fewer items being changed during analyst review (Evans, forthcoming).

4.0 Bureau of Labor Statistics (BLS) Internet Data Collection Experiences

The Bureau of Labor Statistics (BLS) is responsible for collecting information on labor economics. BLS is organized into four program areas plus support offices.

Surveys directly fielded by BLS are establishment based. Collection methodologies vary by program. Federal-State cooperative programs initially field most schedules by mail with telephone follow-up for nonresponse. At the other extreme, most schedules in the compensation and price programs are initially fielded by personal interview due to the complexity of the data requirements with mostly mail used for updating data.

Many surveys are longitudinal giving a respondent a copy of at least his most recent response and allowing edits that can take previous responses as inputs. BLS has begun to give respondents the option of recording their responses on the Internet. The Current Employment Statistics Survey, Annual Refiling Survey, Multiple Worksite Report Survey, Occupational Employment Statistics Program, National Compensation Survey, the International Price Program, the

Producer Price Program and the Occupational Safety and Health Survey all have or will shortly have capability to allow respondents to submit reports via the Internet. This type of arrangement then raises the question of what type and how many edits do we put in place to ensure that we do not have to call back the respondent with questions.

The BLS approach to Internet data collection is to provide a singular, manageable, secure architecture with the following characteristics:

- Same entry point for all BLS surveys
- Common look and feel across surveys
- Support for multi-survey respondents
- Multiple levels of security
- System behind its own firewall, outside of BLS firewall
- Access controlled by BLS-issued digital certificates and passwords
- Monitoring and risk assessment of only one infrastructure.

Here we detail the results of efforts in three BLS surveys to illustrate editing problems associated with different levels of complexity in data-collection requirements. Most BLS surveys are not mandatory. The Occupational Safety and Health Survey (OSHS) is the only mandatory BLS survey.

4.1 Current Employment Statistics Survey (CES)

BLS has explored various mediums to capture the data that makes it as easy as possible for a respondent to participate. The CES has employed touch tone data entry for years to ensure timely capture of the data as efficiently and low cost as possible. CES has arrangements with large responders to receive electronic files containing the required data elements, generally in a standard format. The process greatly reduces respondent burden. Since 1996, CES has had a small portion of its sample report via the Internet.

The CES is a monthly survey of employment, payroll, and hours. The data are used to produce the establishment employment statistics used in the employment situation release, which is a key economic indicator. The sample of 300,000 business establishments provides data that are published after only two and a half weeks of collection. Details about unemployment by industry, state, and area, average hourly earnings and average weekly hours are published. Respondents are queried about five basic items:

- All employees
- Women employees
- Production or nonsupervisory workers
- Production/Nonsupervisory payroll
- Production/Nonsupervisory hours
- Commissions collected for the Service Providing industry
- Overtime hours collected in Manufacturing.

The CES has moved away from mail collection over the last dozen or so years. In use since 1986, telephone data entry (TDE) is the backbone of CES collection. 150,000 reports are collected by TDE. Respondents call a toll-free number and receive a computerized interview. Growth of the Internet and improvement of computer technology allows use of new approaches: Web, Electronic Data Interchange (EDI), and Fax Optical Character Recognition.

CES began collecting data on-line from the web in 1996. One aspect of web methodology is on-line editing—the visual interface allows interaction with respondent not available in other self-administered collection methods such as FAX or mail. TDE systems could edit data and repeat failures back to respondent, however, CES program staff believe the cognitive burden resulting from such edits would be excessive.

Initially the CES Web system performed basic edits only (April '97):

- Logic errors--e.g. All Employment greater than Production Workers
- Range checks--Average hourly earnings between $1.00 - $150.00
- Validity checks--numeric entry, mandatory fields (All employees) completed
- Data entry errors.

The results from these initial basic edits were as follows:

- Approximately 40 percent of current web sample has failed at least one edit check
- Approximately 3 percent of web reports fail one edit check each month
- In 88 percent of all edit failures, the respondent corrected and submitted data during the same session.

After a respondent has entered all of the data elements, he/she submits the report. Once the respondent hits the submit button, the edits are performed. If there is a failure, an edit box appears. The edit messages show as highlighted text below the data section on the instrument (see Appendix B for screen shots). The respondents have to reconcile the error or enter an appropriate comment code and resubmit. For the two global edits, the data must be actually corrected to submit the report, e.g. number of women workers cannot exceed total employment and number of production/nonsupervisory workers cannot exceed total employment.

Because data can be quite variable from month to month, the program managers decided that most edits needed to be soft. Experience with analyst review of edits tends to suggest that most records that fail current machine edits are accepted by the analyst and ultimately used in estimation. Therefore, the only hard edits relate to "impossible" type data situations or entry of nonnumeric values in the data fields.

The current CES Internet application was enhanced in May 2001 to include an expanded array of edits:

- Expanded basic and longitudinal edits
- Critical values vary by major industry division
- Over-the-month changes compared for each data element and for several "calculated" averages.

Up to 21 separate edit checks are performed for each report (depending on the data items on the report form). These edits are performed on the server side.

Except for the two "hard" edits cited above, respondents must either correct so the edit condition is removed or select "comment code" from a drop-down list of common reasons for large fluctuations in the data.

The enhanced edits were patterned after the edit parameters currently used in the CES CATI system. The primary difference was the level of industry detail at which the parameters were set. In the CATI system, the edits were specific based on 2-digit SIC and size of firm. For the Web edits, the parameters were set at the major division level (with size of firm also taken into account).

Experience with enhanced edit in CES Web application: 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 since many additional edit checks were now in place. The percent of units entering a comment code likewise rose from about 6 percent to about 14 percent. By comparison the proportion of TDE self-reported units that enter a comment code is only 3 percent. During CATI interviews the percent of records with a comment code is about 12 percent. So it appears that Web reporters are self-reporting comment codes at about the same rate as interviewers are entering them to explain large fluctuations in the data and considerably more often than touchtone respondent self-report a comment code.

Respondent reaction to longitudinal edits: Our overall experience has been favorable. We provided no advance warning to Web respondents when we introduced the expanded edits. Nevertheless, we had no negative feedback when they were introduced. To further check on this, we 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.

CES staff believe that more research on the cognitive aspects of edits--how they are shown, what wording is used, what options are presented--would be valuable.

4.2 Occupational Safety and Health Survey (OSHS)

The OSHS is a mandatory survey designed to yield detailed specific industrial incidence rates of workplace injuries and illnesses. The survey provides information annually on the number and frequency of nonfatal injuries and illnesses occurring in the workplace. It also reports on the demographics and case characteristics for the serious incidents, those that require time away from work.

The OSHS statistical system is built on the collection of recordable cases. A recordable case is any occupational death, regardless of the time between the injury and death or the length of the illness or nonfatal occupational illness, or nonfatal occupational injury which involves one or more of the following: loss of consciousness, restriction of motion, transfer to another job, or

medical treatment (other than first aid). The OSH Act required that employers subject to the act maintain a standard set of records covering their injury and illness experience. BLS then asks for reports on the summary of incidents at a sampled site with further sampling of individual cases for the event, source, nature, and part of body affected.

In the OSHS production data collection systems, we perform nearly 170 edits (167 in the 2002 system). The edits fall into the categories of validation edits (data are in the correct format), consistency (comparing data between data elements - does it make sense), and reasonableness (primarily for coding - can you really have a funeral director working in a furniture store?).

For the 2002 OSHS survey just fielded, each traditional schedule in 47 States includes a URL and password for a respondent to enter his data directly via the Internet. With this first release of an on-line Internet instrument, we didn't want to overwhelm respondents with many error messages. We made a conscious decision to keep the editing to a minimum flagging only invalid data that would cause the database to crash (where it mattered most) and get the respondents in and out quickly. The edits included in the Internet instrument ensure validity of the data, for example, the value entered in Annual Average Employment must be numeric and the value entered in the Total Number of Cases with Days Away from Work must be numeric.

We expected that we would have to make about the same number of calls to the respondents as we do currently. The edits occur when the respondent presses SAVE (which they can do at anytime on any page) or they press CONTINUE to go to the next page.

The processing system edits establishment summary data for differences in total reported employment from BLS records, data out of range and injury and illness counts unusual for the reported industry. The sampled information requested for individual injuries or illnesses that meet reporting threshold is edited for consistency between the nature and source of the injury versus injury outcome. For example, if the nature of an injury is a sprain the body part affected cannot be the brain or skull.

4.3 National Compensation Survey (NCS)

The NCS is used to collect compensation data. Products derived from NCS data include the Employment Cost Index, locality and national occupational wage levels, national wage and benefit cost levels, benefit incidence, and benefit provisions. These products provide the following information:

- Wages by occupation and work level, for localities, broad geographic regions, and the entire United States
- Employer costs for wages and benefits, and the rate of change in those costs the Employment Cost Index
- Percent of workers receiving benefits and provisions of those benefits.

The NCS samples about 40,000 establishments over 154 PSUs. Within each establishment, a probability-proportionate-to-size sample of occupations is selected. Each occupation is

classified by work level. Wage data are collected in all sampled occupations within the establishments. Benefit data are collected in about 40 percent of the establishments.

Establishments are in the sample for approximately five years with 20 percent replaced each year. For about 60 percent of the establishments, only wage data are collected; with updating occurring once a year. For the remaining establishments, wage and benefit data are collected quarterly. Data are collected on establishment characteristics, occupation characteristics including work level and wages with incentive payments, and employee benefits

First time establishments receive a personal visit from a BLS field economist. Subsequent contacts are via personal visit, mail, phone, or electronic data transmission. Field economists enter data into a laptop-based data capture system developed by BLS. The system can accept data from respondents electronically.

The NCS is developing software to permit the collection of wage data via the Internet. The Internet will be used only to update information previously collected. The addition of Internet reporting will begin on a test basis in the summer of 2003.

The software under development will have seven major features:

- Selected schedules are made available to respondents through the Internet
- Respondents see information from their previous schedules
- Respondents enter updated information
- Basic edits help respondents identify problem situations and clarify or correct data
- Directions and a full-featured help system guide respondents
- Data moves from Internet collection vehicle to the NCS Integrated Data Capture system (IDC) by the field economist
- Field economists import the data into the IDC schedule using standard techniques.

Edit messages will pop up and require respondents to double-check their input. The respondent will have a choice among the following actions:

- Fixing the failed data
- Canceling the changes
- Saving the changes anyway (system flags cells)
- Documenting the situation in a remarks box
- Accessing a help screen.

The initial NCS Internet Collection System will only edit for invalid entries. Warning messages will also pop up for questionable work schedule entries or missing text in text fields.

Edits range from invalid data field entries to negative values for current employment levels, wage rates, and hours worked. For numeric entries, valid values are 0 to 9.

Warning messages are displayed for the following situations:

- The average salary in the occupation has decreased
- The number of workers in the occupation has changed by more than ten percent
- The average salary has increased by more than ten percent
- The work schedule is less than one hour per day or more than 8 hours per day
- The work schedule is less than 32 hours per week for a full time occupation, greater than 40 hours per week for a full-time occupation, or is greater than 32 hours per week for a part-time occupation
- The Current Reference Date is more than 60 days from today
- New employment differs from previous employment by more than 20 percent
- A text field has missing data such as Company Name is missing

Once the respondent completes data entry, the information will be transmitted to a BLS field economist who will have responsibility for reviewing the data and correcting or documenting any additional anomalous data and loading it into the NCS database. In validating the input, the field economist will have the assistance of the full range of data edits that are built into the NCS data capture system. The field economist may need to contact the respondent via e-mail or telephone to resolve any discrepancies that are not explained by documentation. Data edits in the processing system check for occupation code against salary range coded, wage rate compared to the expectation for the local market area, salary range expected for the industry code associated, and so forth. At estimation, outliers are reviewed for validity.

4.4. Summary of BLS Experiences—What We've Learned

BLS began Internet data collection in 1996 with the CES survey. Now at least eight surveys have or plan to have Internet data collection facility in the very near future. Initiation of Internet data collection was targeted to specific establishments while most initial efforts on editing have been on ad hoc basis. Similar themes to the Census Bureau emerge from BLS experience:

- Edits start out at a basic level not to discourage respondent participation. Initial editing ensures that the data submitted are valid (i.e. the numeric data), that mandatory fields have been entered (i.e. the minimum number of fields to ensure a valid response), some gross checks for impossible types of entries
- Edits on Internet applications are kept to a minimum to ensure respondents continue to cooperate
- Edits on the Internet applications are borrowed from post-collection edits
- Program managers prefer to accept data with unresolved errors rather than receiving no data at all
- Reaction of respondents to edits is not shared across the organization.

CES has expanded its edit list over time as the managers have become comfortable with respondents' reactions. OSHS has made Internet data collection an option for most respondents for the 2002 survey. Now is the time to look back at the issues so that we can develop more sophisticated systems that will ensure data quality at reduced cost while maintaining respondent cooperation levels.

5.0 Discussion Points for the Committee

As you can see from the case studies, our agencies have included edits in CSAQ and Internet data collection using a conservative philosophy--at a minimum, we receive data, which may or may or not pass basic edits, from cooperative respondents. We seek discussion on the pros and cons of a more ambitious philosophy, in which we start moving the extensive post-collection edits into Internet data collection in order to reduce costs and increase quality, while maintaining respondent cooperation.

Respondent behavior issues. The following are some issues associated with maintaining respondent cooperation:

- How many edits should there be in a web instrument, relative to the number of questions? At what point does the respondent consider there to be too many edits?

- How should edit information be presented to respondents?

- To what extent can CATI/CAPI edits and post-data collection edits be incorporated into web-based instruments? Which ones? What kinds?

- How complex can CSAQ and Internet edits be? What criteria should be used to judge this?

- How should we word error messages involving variables that are calculated by the system and not entered by the respondent? What types of tools are needed to effectively deliver complex edit-failure feedback information to respondents?

- Are certain types of edits inappropriate for web instruments? What types of limitations should be considered?

- What types of information should we collect during cognitive pretesting and usability testing to guide the types of edits incorporated into CSAQs as well as how well edit-failure information is being communicated?

- If respondents need additional help to resolve one or more edit failures, how should it be provided?

- Will respondents react negatively to being asked to resolve data-collection edit failures and also receiving a clarification call from the Agency?

Data quality issues. Issues concerning data quality and resource allocation can arise when large mail surveys change to or offer the option of web-based data collection. These issues arise from the fact that large mail surveys have high variable costs (with respect to number of respondents and number of survey cycles) associated with data editing because clerks and subject-matter

experts review the edit information produced by post-data-collection edits. Editing at the time of data collection, on the other hand, by the respondents' reviewing edit information from web-instrument edits can have high fixed costs for programming and questionnaire testing, but the corresponding variable costs associated with data editing should be much lower than those for post-data-collection editing. The following are some of the issues associated with this situation:

- Can some (or all) post-data-collection edits be eliminated by incorporating them into the web instrument?

- When post-data collection edits are incorporated into a web instrument, what critical values should be used—same as those used in CATI or post-data-collection systems or different values?

- Will there be an "editing mode" effect—that is, different responses to edits in a self-administered instrument vs. to traditional call backs requesting clarification?

- How should the effectiveness of editing be measured, permitting comparisons of different modes of editing?

- What are the consequences of the current strategy favoring measurement error over nonresponse error? What criteria should be used to determine the optimum tradeoff?

- How does one prevent overediting, such as editing out a trend?

- To what extent will editing at the time of web-based data collection reduce the number of failed edits occurring in post-data-collection editing?

- If web-collected data contains fewer edit failures than mail-returned data, can survey resources devoted to post-data-collection editing be reduced without degrading data quality?

- Will BLS and the Census Bureau be able to publish estimates earlier because of earlier receipt of data and less time spent reviewing post-data-collection editing information?

- What types of research should be conducted to determine the effects of CSAQ data collection on data quality?

Overall strategy issues. Although survey practitioners would very much like to have "generally accepted practices" or "rules of thumb" for many of these design issues, we expect this to be virtually impossible given the variety of surveys, administrations, and trade offs related to data quality and response. Instead we think it would be more appropriate to develop a set of guidelines to aid decisions related to editing.

- What criteria should guide the determination of editing strategies in web surveys?

- What strategy should be employed in evaluating overall editing performance? How should one trade off edits in CSAQs vs. post-collection processing?

- How should one prioritize edits in selecting the edits to be included in a CSAQ?

- How can one transition new surveys to the web without implementing web edits in stages as has been the philosophy to date?

- What types of empirical analyses of edit results (CSAQ and post-collection) should be conducted to provide feedback into the improvement of questionnaires (so that respondent errors may be prevented in the first place)? What are some current experiences in this area and the type of benefits that may be realized?

- Are there sources for working papers on studies not in the published literature on Internet data collection that address data editing? Do committee members know what the private sector and/or European community doing in this area?

We would appreciate feedback from the committee on these issues, as well as guidance based on members' experience or familiarity with incorporating edits into CSAQs and web surveys. We are particularly interested in what the committee believes are the key supporting research questions in this area.

6. Acknowledgements:

The authors want to thank the following staff for their help in gathering the background information for this paper:

Philip Doyle, BLS
William McCarthy, BLS
Dee McCarthy, BLS
Richard Rosen, BLS
Patrick Kent, Census Bureau
Joyce Kiessling, Census Bureau
Yvette Moore, Census Bureau
John Nogle, Census Bureau
Yolando St. George, Census Bureau
Rita Williamson, Census Bureau

7. References

Bzostek, J. and Mingay, D. 2001. "Report on First Round of Usability Testing of the Private School Survey on the Web." Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #42.

Dumas, J. and Redish, J. 1999. *A Practical Guide to Usability Testing*. Portland, OR: intellect.

Fellegi, I.p and Holt, D (1976). "A Systematic Approach to Automatic Edit and Imputation." *The Journal of the American Statistical Association*, 71, pp. 17-35.

Economic Electronic Style Guide Team. September 28, 2001. "Style Guide for the 2002 Economic Census Electronic Forms" U.S. Census Bureau, Economic Planning and Coordination Division.

Evans, E. Forthcoming. "QFR CSAQ Evaluation." Internal Memorandum. U.S. Census Bureau, Company Statistics Division.

Nichols, E. 1998. "Results from usability testing of the 1998 Report of Organization CSAQ" Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #19.

Nichols, E., Murphy, E, and Anderson, A. 2001a. "Report from Cognitive and Usability Testing of Edit Messages for the 2002 Economic Census (First Round)" Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #39.

Nichols, E., Murphy, E, and Anderson, A. 2001b. "Usability Testing Results of the 2002 Economic Census Prototype RT-44401" Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #49.

Nichols, E., Saner, L, and Anderson, A. 2000. "Usability Testing of the May 23, 2000 QFR-CSAQ (Quarterly Financial Report Computerized Self-Administered Questionnaire)" Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #33.

Nichols, E., Tedesco, H., King, R., Zukerberg, A, and Cooper, C. 1998. "Results from Usability Testing of Possible Electronic Questionnaires for the 1998 Library Media Center Public School Questionnaire Field Test." Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #20.

Rao, G. and Hoffman, R. 1999. "Report on Usability Testing of Census Bureau's M3 Web-Based Survey" Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #26.

Saner, L., Marquis, K., and Murphy B. 2000. "Annual Survey of Manufacturers Usability Testing of Computerized Self-Administered Questionnaire Findings and Recommendations Final Report" Internal Memorandum. U.S. Census Bureau, Statistical Research Division, Usability Lab, Human-Computer Interaction Memorandum Series #30.

Sigman, Richard (2001). "Editing and Imputation in a Standard Economic Processing System," *Proceedings of Statistics Canada Symposium 2001*, Ottowa, Ontario: Statistics Canada.

Stinson, Linda L. and Fisher, Sylvia K. 1996. Overview of Data Editing Procedures in Surveys Administered by the Bureau of Labor Statistics: Procedures and Implications, Presented at the First International Computer-Assisted System Information Computing Conference at San Antonio, Texas.

Sweet, E. and Ramos, M. 1995. "Evaluation Results from a Pilot Test of a Computerized Self-Administered Questionnaire (CSAQ) for the 1994 Industrial Research and Development (R&D) Survey," Internal Memorandum. U.S. Census Bureau, Economic Statistical Methods and Programming Division, #ESM-9503.

Tedesco, H., Zukerberg, A., and Nichols, E. 1999. "Designing Surveys for the Next Millennium: Web-based Questionnaire Design Issues," *Proceedings of the Third ASC International Conference*. The University of Edinburgh, Scotland, UK, 22nd-24th of September 1999, pp. 103-112.

Wagner, Dennis L. (2002). "Economic Census General Editing – Plain Vanilla," *Proceedings of the International Conference on Establishment Surveys*, Alexandria, VA: American Statistical Association, pp. 561-570.

Appendix A – Editing at BLS

Stages of editing

Most systems at BLS allow editing to occur throughout the survey process—from collection through the estimation steps. Editing occurs at:

- Interviewer Level
- Regional Office or State Level
- National Office Level
  - Pre-estimation
  - Post-estimation.

BLS Editing system allows manual review and changes:

- At data entry, eg while reviewing paper schedules or notes
- During data entry
- After data entry.

Editing software:

- Requires substantial data cleaning during data editing process
- Requires manual fixes of machine-generated signals
- Has batch capability with manual resolution of errors.

Characteristics of various BLS Editing systems include:Data entry and data edit at the same time (usually performed only on sub-sections of data)

- Data to be entered with correction (high speed)
- Full-screen data entry
- Verification (i.e., enter data twice)
- Implied decimals to be typed (e.g., type 23 for 2.3)
- Data-entry statistics to be generated.

In order to facilitate data analysts, editing systems have various reporting features:

- Lists of missing reports
- Reports
- Log or trace files
- Tables
- Use of Logical operators
- External; Records.

Most editing systems in use at BLS have many features in common. For example, most editing systems have the capability to edit continuous data, decimal values, character data, and some binary checks, such as would be used for categorical data. Some systems imbedded in the data entry process itself and include many of the typical CATI/CAPI editing features, such as the capacity to enter and edit data at the same time through the pre-programmed specification.

- Alphabetic or numeric characters

- Acceptable numeric ranges
- Consistency checks (e.g. marital status versus age)
- Routes and skip patterns
- External standards for comparison.
- Decimal in wrong location
- Wrong Yes/No choice
- Wrong response option
- Wrong measurement unit (week, day, month, year)
- Wrong numeric entry
- Changing decimal locations.

By contrast, some errors many not necessitate editing and are unlikely to have an impact upon data quality and accuracy. A common example of this type of edit would be the presence of typos that do not obscure the meaning of text

Systems employ a variety of edit types to check for logical relationships within the schedule and across the historical database. These types of edits include

- Linear edits (i.e., their graphs are straight lines)
- Non-linear edits
- Conditional edits
- Ratio edits
- Variables that can accept negative values (negative values are edited out)
- Quantitative edits
- Qualitative edits
- Univariate range checks
- Multivariate, record-level checks for consistency
- Other consistency checks (e.g., work schedule if full-time employee)
- Route checks or skip patterns
- Check of generated values against external standards.

Editing systems are designed to:

- System can accept logical operators (e.g., 'and,' 'or,' 'not')
- System can perform intra-record checks
- States, regions etc. have customized error limits
- Software performs graphical inspection of the data.

Editing systems can perform the following functions:

- Check edits for redundancy
- Check that edits are not contradictory
- Generate implied edits
- Generate external records.

System can perform statistical edits based upon:

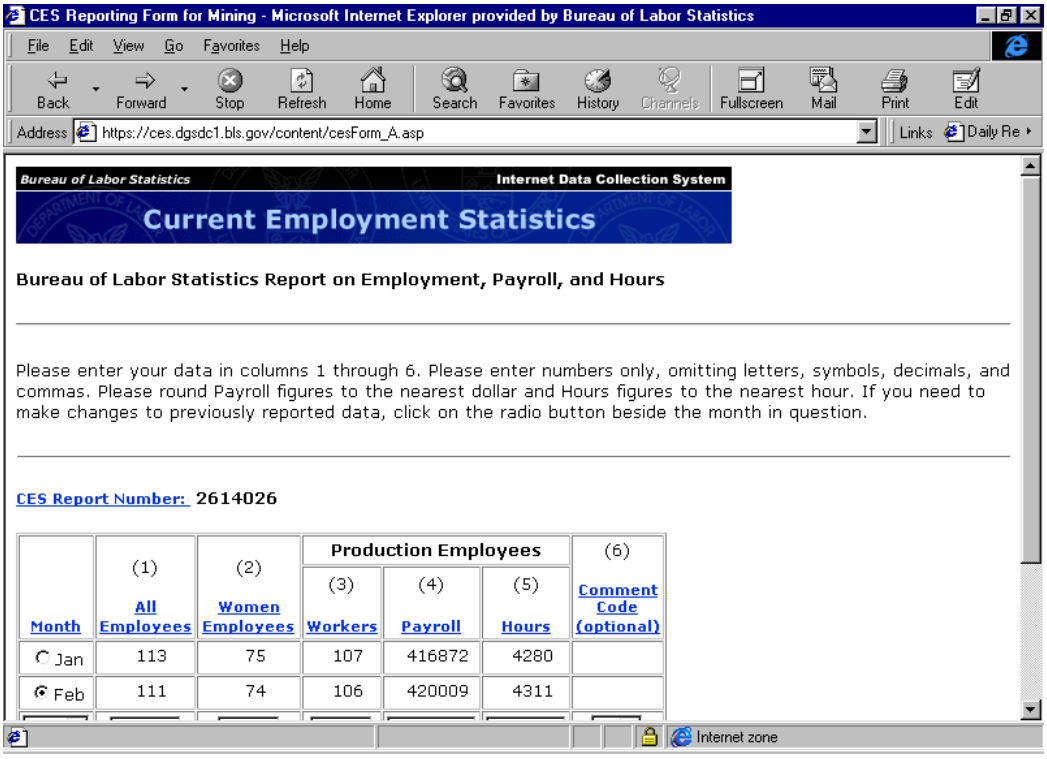
- Historical data

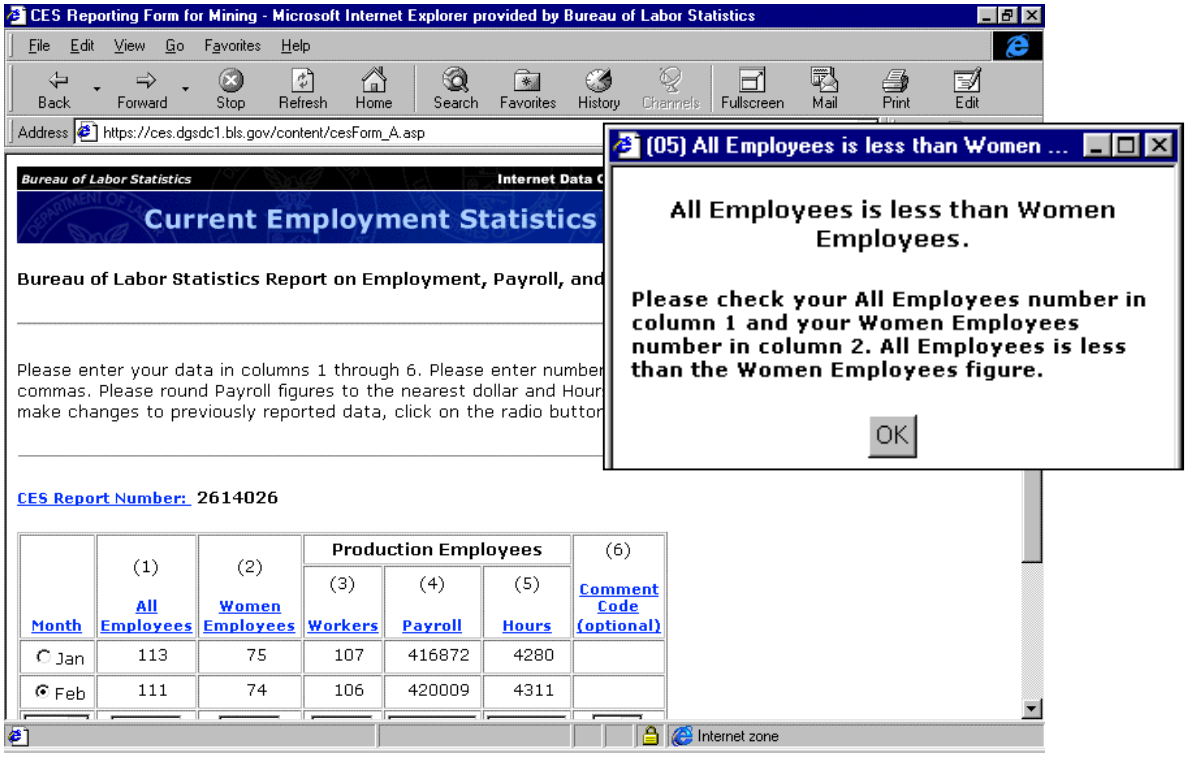
- Cross-record checks
- Univariate outlier detection
- Complexity of data structure.

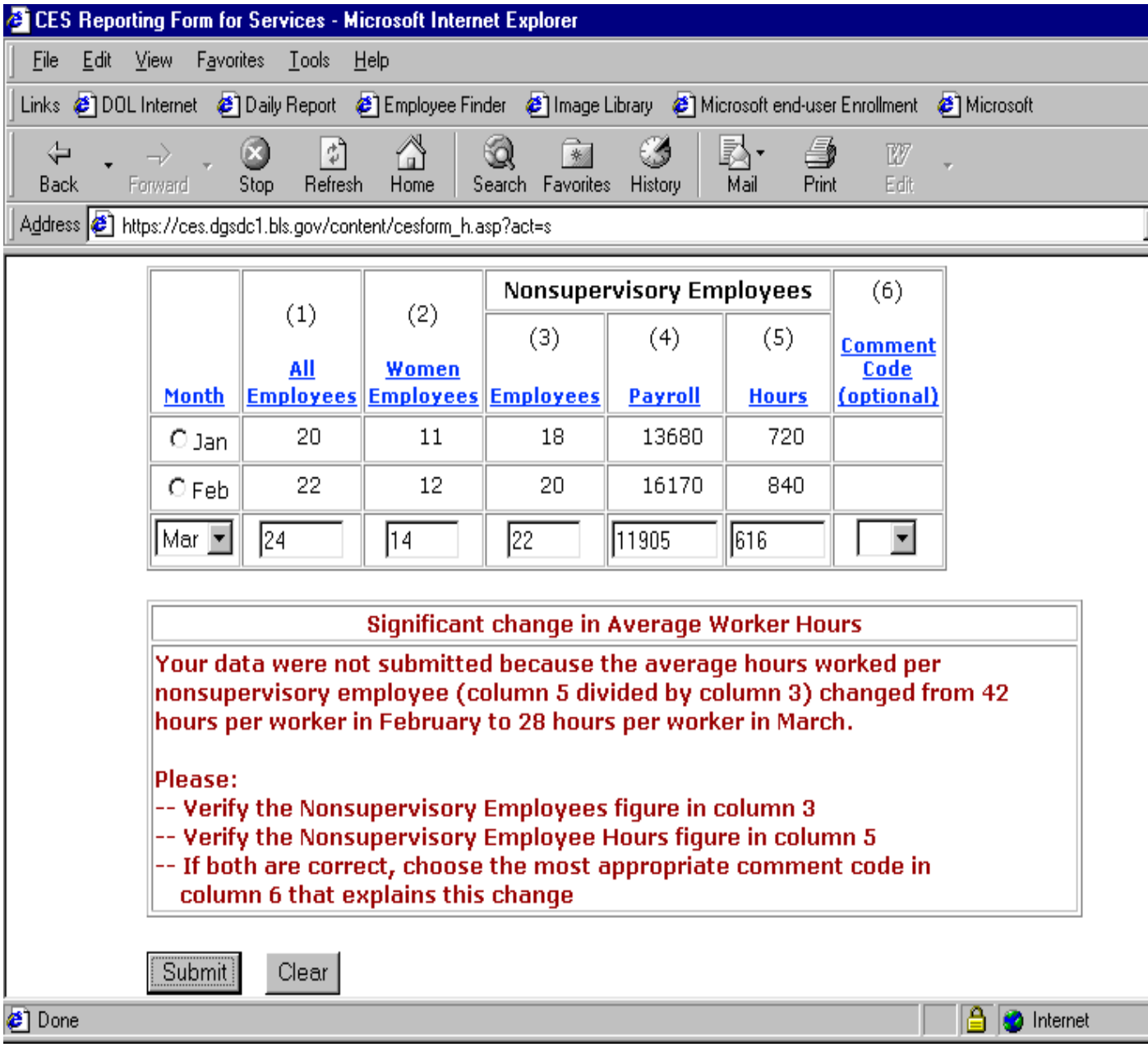
Editing system allows:Manipulation of hierarchical data

- Manipulation of complicated skip patterns
- Manipulation of subfiles
- Manipulation of cross-record checks
- Respecification of edits without other manual software changes
- Respecification of data without other manual software changes.

Appendix B CES Edit Screen Shots




CES Reporting Form for Mining - Microsoft Internet Explorer provided by Bureau of Labor Statistics
File Edit View Go Favorites Help
Back Forward Stop Refresh Home Search Favorites History Channels Fullscreen Mail Print Edit
Address https://ces.dgsdc1.bls.gov/content/cesForm_A.asp
Bureau of Labor Statistics
Internet Data C
Current Employment Statistics
Bureau of Labor Statistics Report on Employment, Payroll, and
Please enter your data in columns 1 through 6. Please enter numbe
commas. Please round Payroll figures to the nearest dollar and Hour
make changes to previously reported data, click on the radio buttor
CES Report Number: 2614026
Month
(1) All Employees
(2) Women Employees
Production Employees
(3) Workers
(4) Payroll
(5) Hours
(6) Comment Code (optional)
Jan 113 75 107 416872 4280
Feb 111 74 106 420009 4311
Internet zone
(05) All Employees is less than Women ...
All Employees is less than Women Employees.
Please check your All Employees number in column 1 and your Women Employees number in column 2. All Employees is less than the Women Employees figure.
OK


CES Reporting Form for Services - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Links DOL Internet Daily Report Employee Finder Image Library Microsoft end-user Enrollment Microsoft
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit
Address https://ces.dgsdc1.bls.gov/content/cesform_h.asp?act=s
Nonsupervisory Employees
(1) All Employees
(2) Women Employees
(3) Employees
(4) Payroll
(5) Hours
(6) Comment Code (optional)
Month
Jan 20 11 18 13680 720
Feb 22 12 20 16170 840
Mar 24 14 22 11905 616
Significant change in Average Worker Hours
Your data were not submitted because the average hours worked per nonsupervisory employee (column 5 divided by column 3) changed from 42 hours per worker in February to 28 hours per worker in March.
Please:
-- Verify the Nonsupervisory Employees figure in column 3
-- Verify the Nonsupervisory Employee Hours figure in column 5
-- If both are correct, choose the most appropriate comment code in column 6 that explains this change
Submit
Clear
Done
Internet