

OVERVIEW OF DATA EDITING PROCEDURES IN SURVEYS

ADMINISTERED BY THE BUREAU OF LABOR STATISTICS:

PROCEDURES AND IMPLICATIONS

Linda L. Stinson
Research Psychologist
Bureau of Labor Statistics
2 Massachusetts Avenue, N.E.
PSB Room 4915
Washington, D.C., 20212

Sylvia Kay Fisher
Research Psychologist
Bureau of Labor Statistics
2 Massachusetts Avenue, N.E.
PSB Room 4915
Washington, D.C., 20212

A paper to be presented at the first International Computer-Assisted System Information Computing (CASIC) Conference to be held in San Antonio, Texas on December 12, 1996.



ABSTRACT

The purpose of this study is to identify issues associated with cleaning computerized data files, using the example of large-scale data collection activities conducted by the Bureau of Labor Statistics (BLS). This paper will describe data editing activities and processes currently implemented to clean data collected by BLS through CAPI and CATI interviews. Similarities and differences in current data editing procedures are described, as well as common and unique decision rules used by the various BLS surveys in the conduct of both household and establishment surveys.

This study describes features associated with data editing software systems in place at the BLS; identifies data editing issues shared in common among BLS surveys; and documents the types of data editing activities and procedures currently implemented at the BLS, as well as how these procedures address data editing needs. This integrated profile is designed to provide an overview of major data editing activities conducted by the BLS to improve data quality that can enhance and inform data reporting.

OVERVIEW OF DATA EDITING PROCEDURES IN SURVEYS ADMINISTERED BY THE BUREAU OF LABOR STATISTICS: PROCEDURES AND IMPLICATIONS

INTRODUCTION AND PURPOSE OF STUDY

Researchers engaged in the administration of large-scale regionally- and nationally-administered surveys encounter numerous difficulties when processing incoming data. Many problems in data collection can impair and affect the quality of data generated from these surveys. Cleaning up data is an imperative aspect of enhancing data quality for statistical analysis purposes. The process of cleaning up data to ensure and verify its accuracy is called data editing.

Currently, data editing is an under-described, albeit immensely important, component of the data collection process. Data editing processes and procedures are governed by an interaction between available computer technologies and decision-making ‘rules,’ themselves often a function of the idiosyncrasies of the survey instrument. While effective and appropriate data editing procedures are designed to enhance the quality of survey data and minimize error, many issues must be weighed when selecting and developing an effective data editing system that promotes data quality while effectively utilizing available resources.

The purpose of this study is to identify issues associated with cleaning up survey data, using the example of large-scale data collection activities conducted by the Bureau of Labor Statistics (BLS). This study will provide an overview of field data activities undertaken by Federal agencies conducting CAPI- and CATI-interviewing survey projects, describe both common and idiosyncratic decision rules used in field data edits, and identify common processes undertaken by these agencies. This paper will describe those data editing activities and processes currently being implemented to ‘clean data’ collected by BLS. Little research has been conducted to describe issues associated with identifying data editing needs, data editing procedures and the decision-making criteria that should be applied to resolve data editing issues. This paper will document similarities and differences in currently applied data editing procedures, describe both common and idiosyncratic decision rules used in field data edits, and identify common data editing processes conducted across household and establishment surveys.

This study will describe each of the following:

1. Features associated with data editing software systems in place at the BLS;
2. Identify data editing issues shared in common among BLS surveys;
3. Types of data editing activities and procedures are currently implemented at the BLS, and how they address data editing needs; and

4. Levels at which data collection activities are undertaken and implementation activities at various levels of data editing.

BACKGROUND

Human error, such as programming mistakes, miscalculations, key punch 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. But in all cases, the goal is the same: to identify and correct as much error as possible.

In April, 1990, the Federal Committee on Statistical Methodology reported in the “Statistical Policy Working Paper 18 - Data Editing in Federal Statistical Agencies” that Federal survey managers consider data editing an important component of the survey process. The Subcommittee used the following working definition of data editing to conduct their analysis of federal-level data editing procedures:

Procedure(s) designed and used for detecting erroneous and/or questionable survey data (survey response data or identification type data) with the goal of correcting (manually and/or via electronic means) as much of the erroneous data (not necessarily all of the questioned data) as possible, usually prior to data imputation and summary procedures.

To that end, 60% of those managers engaged subject matter specialists and editors for the express purposes of review and resolution. Over 80% of their surveys maintained good internal documentation of the editing systems they used (p. 2). But it was recognized that we stand on the verge of daily technological advances; there is substantial potential for new advances in data editing through the use of expert systems, computerized data collection, and artificial intelligence. The Committee, therefore, recommended that Federal agencies should:

1. Evaluate and examine the cost efficiency, timeliness, productivity, repeatability, statistical defensibility, and accuracy of their current editing practices;
2. Continue to share information on research and development of software systems for editing;

3. Evaluate the role and effectiveness of editing in reducing nonsampling errors for their surveys;
4. Evaluate the relationship of extensive manual review on resulting estimates; and.
5. Explore integration of data entry and data editing function. (pp. 3-4, Working paper 18 citation)

Taking heed of the Committee's advice, this paper begins the first step in the long process of investigating editing procedures at the Bureau of Labor Statistics (BLS). At this preliminary stage, we present an introduction and overview of the editing that is currently being done on the major BLS surveys. We will look at both the micro- and macro-levels of data editing.

Typical Editing Activities: Many types of edits are performed with large scale survey data collection projects. One type of editing comprises *immediate* edits, including activities such as typing corrections on open-ended CAPI-CATI questions, and automatic edit checks when out of range answers are not accepted, a feature typically found with CATI/CAPI systems (Couper, personal communication, 2/5/96).

A second type of edit is further removed, and entails supervisor editors who use several means (including the use of recognizable blue pencil) to identify out of range responses, incorrect skip patters etc. Central office editors (like the lower level economists at BLS) often conduct additional edit activities, by looking for substantively incorrect/impossible answers, as well as comparability with previous answers; this is particularly true of large panel studies, where comparability is an important issue.

Types Of Errors That Require Editing: Certain types of edits are necessary to ensure data quality and accuracy. Examples of this type of edit include:

- decimal in wrong location
- faulty skips leading to wrong path
- wrong Yes/No choice
- wrong response option
- wrong measurement unit (week, day, month, year)
- wrong numeric entry
- (NEW) 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

Zapf et al. (1992) has distinguished three levels of action regulation that characterize typical errors found in data. That is, many data errors stem from one of these three sources:

- INTELLECTUAL LEVEL (thought, memory, and judgment errors)
- FLEXIBLE ACTION PATTERNS (errors of habit, omission, or recognition)
- SENSORIMOTOR LEVEL (e.g., typographical errors)

Zapf et al. further point out that the majority of CATI errors occur at the SENSORIMOTOR level. Couper, Hansen, Sadosky (1996, p. 6) have identified the types of errors that stem from this typology; they have also specified how these errors can impact upon the results, because of their ability to obscure or reduce accuracy.

In the interview situation, editing may be done at three stages. First of all, there are the immediate edits done by interviewers. In the CATI/CAPI environment, these may include (1) correcting typographical errors and (2) changing responses that fall out of range and are not accepted by the computerized instrument. In the paper/pencil (PAPI) interview, this immediate editing may include re-writing handwritten notes for purposes of legibility. Secondly, editing may be done by supervisors. This is generally done for PAPI and not for CATI/CAPI. In this case, supervisors check for out-of-range response and/or incorrect skip patterns. Thirdly and again for PAPI, editing may be done by the central office. At BLS, this includes editing by economists who look for substantively incorrect or impossible answers or comparability with previous answers when it is a panel study.

METHODOLOGY

Data collection activities for this study focused on the conduct of extensive interviews with BLS personnel responsible for various dimensions of the data editing process. To ensure representativeness, experts were selected for each of four major survey projects implemented by BLS, including:

1. The SPAM System used to edit the Occupational Employment Survey (OES),
2. The IDC Data Capture System used to edit the new COMP 2000 survey
3. The ARIES System used to edit the Current Employment Survey
4. The ECI Data Entry and Data Editing system

These four survey projects were also chosen, because many other data editing projects are currently undergoing major changes to either update and improve the existing system or replace the existing system with a new data editing system.

Interviews focused on discussion of data editing procedures implemented at the local, regional, and national sites, as well as every point of the data collection and analysis process at which data editing occurs. This aspect of interviews focused on the types of edits applied during data collection, data transmission, data analysis and processing, and post-analysis data treatment.

A second series of interviews was conducted with experts who specialize in the computer programming and technology used for data editing. The survey instrument developed by the Federal Subcommittee on Statistical Methodology was administered to BLS employees responsible for developing and/or using data editing computer software. This instrument examines several aspects of data editing computer software, including: types of edits and statistical imputations performed by software system, graphical features associated with the software, computer features needed to apply the software editing system, language(s) used for

programming and analysis purposes, etc. This instrument also served as the central component of extensive interviews with data editing personnel who specialize in associated data editing software. The Survey of Data Editing Software Features may be found in Appendix I.

Since so much of the BLS data collection is done in a CAPI or CATI mode, questions about computerized editing of data are especially pertinent. To better understand and document BLS computerized editing processes, we turned to the checklist of functions and attributes for editing systems created by the OMB subcommittee on Editing in Federal Agencies (Statistical Policy Working Paper 18, Data Editing in Federal Statistical Agencies, Appendix C), developed to assist agencies in evaluating hardware and software. The description of 4 major editing systems used at BLS will be described here.

General Features: While only one of these, the ARIES system, is actually a CATI editing software, all four share many features in common. For example, each of them have the capability to edit continuous data, decimal values, character data, and some binary checks, such as would be used for categorical data. Three of the systems, ARIES, SPAM and ECI, are incorporated into 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 of:

- alphabetic or numeric characters,
- acceptable numeric ranges,
- consistency checks (e.g. marital status versus age),
- routes and skip patterns, and
- external standards for comparison.

Of the four editing systems, ARIES and SPAM go beyond data entry and are also used as tabulation and analysis editing systems. As such, they function as part of larger software systems, the EMPLOYMENT AND EARNINGS SYSTEM and the CV SYSTEM respectively. Perhaps the greatest value of these editing systems is their capacity for macro edits.

While none of these four editing programs are in the public domain, they have features that should be looked for when selecting an editing system. There are some overarching issues associated with the implementation of data editing techniques and procedures and these issues should be addressed when examining and considering implementing new data editing approaches. Among these issues are the following:

1. General Features
2. Survey Management
3. Systems Items
4. Edit Writing
5. Types of Edits
6. Data Review and Correction
7. Support, Updates, and Training

I. GENERAL FEATURES

This section of the results focuses on the description of the general features associated with BLS data editing systems, particularly emphasizing the primary functions and capabilities available within these data editing systems. In addition, the types of software environments BLS data editing systems use and are embedded within are included here. To begin with, it should be noted that existing BLS data editing systems can treat all types of data, including:

- Categorical data (e.g., 1=female, 2=male)
- Continuous data (a number represents a quantity)
- Decimal values
- Character data (text accepted as data)

As may be seen in Table 1, BLS data editing systems span a full range of capabilities, beginning with editing systems that emphasize data entry and editing processes exclusively, to more extensive and complex systems that allow for CATI and CAPI data collection. Some of the editing systems are designed to yield tabular data and contain summary systems used for preparing tables of summary data. Although some BLS editing systems emphasize data entry and editing features exclusively, the SPAM system has a full complement of features that permit editing as well as imputation procedures to be implemented.

The SPAM software system is also a part of a larger CV software system, and has an extension system available within the larger program office. In the case of ARIES, the editing system is subsumed under the greater Employment and Earnings editing system, which is located within the program office, and permits the use of imputation procedures. The COMP 2000 is similar to SPAM and ARIES, because it is associated with an extensive system available within the larger program office. Although the majority of BLS editing software systems are part of a larger software system, the ECI Data Editing and Data Entry System is a self-contained editing system.

Both the COMP 2000 and SPAM systems utilize a mini-computer to warehouse its editing software, whereas the ARIES and ECI systems necessitate mainframes to house the editing system. It should be noted that the ARIES system is currently being revitalized to eventually be placed upon a super-mini or Sun workstation; it is anticipated that this renovation will be in place within one or two years. This reflects a growing trend that is likely to continue to flourish throughout the BLS, in which mini-computers or super-mini-computer systems will be used to replace existing mainframe systems. The gains to be incurred from this change include the facilitation of access, reduction of storage space, and reduction of the expenses associated with larger mainframe systems.

The editing systems can be used on several operating systems. ECI and SPAM utilize an OS/2 system, whereas the ARIES and COMP 2000 systems operate in an MS DOS system. It is possible that in the near future, the majority of BLS editing software systems will operate within

MS DOS systems as the trend towards placing editing systems currently on mainframes onto smaller higher speed micro-computing environments continues.

The ECI system is currently using a Foxpro test system, but will utilize Powerbuilder in the relatively near future. The COMP 2000 system is currently utilizing Powerbuilder operating system. The ARIES system uses Foxpro, Turbo Pascal, and Excellink, which is the host operating software system. The SPAM system also uses a Foxpro test system, but that may change when the new editing system is in place within the next two to three years.

| TABLE 1 | |
|--|--|
| MAJOR FEATURES OF BLS DATA EDITING SYSTEMS | |
| I. SPAM System: OES | <ul style="list-style-type: none"> • an editing and imputation system • a data entry system • a tabulation system • a data analysis system • a summary system for preparing tables etc. |
| II. ECI Data Editing and Data Entry System | <ul style="list-style-type: none"> • an editing system • a data entry system |
| III. Aries Editing System: Current Employment Survey | <ul style="list-style-type: none"> • a CATI system • a data entry system • a tabulation system • a data analysis system • a summary system for preparing tables etc. |
| IV. COMP 2000 | <ul style="list-style-type: none"> • an editing system • a data entry system • a data analysis system |

Currently, none of the BLS editing software is available within the public domain. However, much of the editing software is available to other programs within the BLS; in addition, the SPAM system has been released to each of the 50 states, specifically the equivalent of the Office of Employment offices in each state's government. Thus, each state officer responsible for providing data to the SPAM system has access to the system directly within the state site.



II. SURVEY MANAGEMENT

This aspect of data editing process focuses on those features within each editing system that are responsible for the management of survey data. In particular, this aspect of the survey examines the types of manipulations and procedures available within each system that serve to track the processing of data records throughout the data input, processing, and cleaning stages. The survey management components of editing systems are important because they document the tracking of survey data throughout the every stage of data processing, which includes the editing process. Survey management ensures the integrity of the database by recording the number of times edit rules fail, creating audit trails, and evaluating whether records are intact and consistent throughout data processing. Examination of Table 2 indicates that several survey management activities are commonly conducted within these BLS editing systems.

| | |
|---|---|
| I. SPAM System: OES | <ul style="list-style-type: none">• Provides check-in records against a sample master• Generates a list of missing reports• Indicates which records are in need of recall or recheck• Determines when extra records are present• Creates an audit trail by tracking all changes made• Provides reports• Determines the number of times each edit rule failed• Provides information on enumerator performance (for imputation purposes)• Determines the number of times donor values were assigned (for hot- or cold-deck imputation procedures) |
| II. ECI Data Editing and Data Entry System | <ul style="list-style-type: none">• Provides check-in records against a sample master• Generates a list of missing reports• Provides data on the cost of the editing process• Indicates which records are in need of recall or recheck• Creates an audit trail by tracking all changes made• Provides reports |

TABLE 2

SURVEY MANAGEMENT ACTIVITIES IN BLS EDITING SYSTEMS (Continued)

| | |
|---|---|
| <p>III. Aries Editing System: Current Employment Survey</p> | <ul style="list-style-type: none"> • Provides check-in records against a sample master • Generates a list of missing reports - specifically key reporters which serve as subsets of data that are examined • Indicates which records are in need of recall or recheck • Determines when extra records are present - examines for duplicate records using identification number • Creates an audit trail by tracking all changes made |
| <p>IV. COMP 2000</p> | <ul style="list-style-type: none"> • Provides check-in records against a sample master • Generates a list of missing reports • Provides data on the cost of the editing process • Indicates which records are in need of recall or recheck • Creates an audit trail by tracking all changes made • Provides reports |

Among those features commonly available within editing software systems are the provision of check-in records against a sample master, the generation of a list of missing reports, an ability to indicate which records are in need of recall or recheck, the provision of reports, and the ability to create an audit trail by tracking all changes made. These features allow the regularity of the flow of data within the system to be tracked, ensure quality control, and allow system managers to ensure that reporters from previous years are still providing reports within the current year. It should also be noted that the SPAM system further contains survey management features associated with the implementation of editing and imputation systems, including information on enumerator performance and the determination of the number of times donor values were assigned through imputation techniques.

The SPAM system also allows for the determination of the number of times each edit rule that failed, which allows users to identify the number of records failing to meet edit standards. Although data editors in other programs favored the collection of this information, some reported that their responsibilities for completing production schedules were too time-consuming to allow for the regular collection of this information. It should also be noted that none of the four systems generates data on the cost of the data editing process.

III. SYSTEMS ITEMS

This component of the editing software questionnaire emphasized items examining the operational features of BLS editing software systems. The first question focused on the types of setups that can be generated by each of the editing software systems. The ECI and SPAM systems are both compiled, rather than interpreted systems, which permits faster execution of treating data. In addition, users don't need a copy of the programming code to use interpreted systems, which facilitates their use in the editing process (OMB, May, 1990). Both the ECI and SPAM editing systems use ASCII-II to generate setups for editing programs and both provide field-by-field editing screens. The COMP 2000 system is also a compiled system that uses SAS to generate setups and which utilizes field-by-field editing screens.

The ARIES system is a compiled system that uses Dbase to generate setups; in addition, both spreadsheet and graphics programs have been developed especially for the ARIES system by BLS personnel. The ARIES system is programmed with preset record layouts and field specifications. These fields are used to identify the record, and also allow breakdowns of records to be specified, including records by geographic regions. The availability of this feature improves the quality of reporting, and allows state-by-state data to be compared or aggregated as needed.

Table 3 provides a summary of the sources of input data that are treated by each of the editing systems. Inspection of this table indicates that the editing systems collect sources of input data from varying sources. As might be expected, the ARIES system collects input data from the largest number of sources, because the CES permits reporters to transmit their information through any means the reporters themselves have available. Because of this flexibility in data collection methods, the ARIES system was especially designed to permit the integration of these varied sources of data within the editing system. By contrast, the other three systems permit the collection of ASCII-II formatted data and/or input data that is provided directly within each system's data entry software.

Table 4 provides an overview of many of the operational features associated with each of the BLS editing systems. The operational features examined here emphasize several aspects of how the editing system is designed, and how these design features impact the editing system's functioning. These features include:

- the manner in which the editing system is designed;
- how storage is constructed within the editing system;
- the ability of the system to "call" or access user-programmed sub-routines;
- the "embedding" of the system within a larger database, and the nature of that database, if appropriate;
- the generation of trace files; and
- the security of the system, as epitomized by the use of passwords and other features.

Certain operational features are consistent within BLS editing systems. One feature is the construction of the editing system in modules. This operational feature allows parts of the editing systems to be updated without affecting other system modules, which serves as a time-saving device. The availability of specialized storage structures serve to facilitate the storage of newly collected data and facilitates the retrieval of subsets of data drawn from the greater database. The

ability to “call” user programmed sub-routines allows for speedy processing of data that must be specially treated. Finally, the availability of passwords or other security measures help ensure the integrity of data throughout the data collection and editing process.

| TABLE 3 | |
|--|---|
| SOURCES OF INPUT DATA FOR BLS EDITING SYSTEMS | |
| I. SPAM System: OES | <p><i>Input data accepted from:</i></p> <ul style="list-style-type: none"> • Data entry software • ASCII formatted data |
| II. ECI Data Editing and Data Entry System | <p><i>Input data accepted from:</i></p> <ul style="list-style-type: none"> • ASCII formatted data |
| III. Aries Editing System: Current Employment Survey | <p><i>Input data accepted from:</i></p> <ul style="list-style-type: none"> • CATI • CAPI • Data entry software • Data files generated from other software • ASCII formatted data |
| IV. COMP 2000 | <p><i>Input data accepted from:</i></p> <ul style="list-style-type: none"> • Data entry software • ASCII formatted data |

All of the BLS editing systems are embedded within a larger database which reflects the information needs for the given survey. For example, the SPAM system is linked to three sources of data, a sampling frame, historical data from previous administrations of the OES survey, and summary-level data from previous administrations. The availability of this feature allows numerous comparisons to be drawn between newly collected data and other data embedded within the database.

| TABLE 4 | |
|--|--|
| OPERATIONAL FEATURES OF BLS EDITING SYSTEMS | |

| | |
|---|---|
| <p>I. SPAM System: OES</p> | <ul style="list-style-type: none"> • System constructed in modules <i>System embedded within a database, with the following features:</i> <ul style="list-style-type: none"> • data from a sampling frame • historical data from previous surveys • previous summary level data |
| <p>II. ECI Data Editing and Data Entry System</p> | <ul style="list-style-type: none"> • System constructed in modules • Software has specialized storage structures • Software allows “calls” to user sub-routines <i>System embedded within a database, with the following features:</i> <ul style="list-style-type: none"> • data from a sampling frame • Software generates a log or trace file of every key stroke (this option is available, but is very expensive, so this editing feature has been turned off) • System uses passwords or other security measures |
| <p>III. Aries Editing System: Current Employment Survey</p> | <ul style="list-style-type: none"> • System constructed in modules • System has specialized storage structures • System allows “calls” to user sub-routines <i>System embedded within a database, with the following features:</i> <ul style="list-style-type: none"> • data from a sampling frame • historical data from previous surveys • previous summary level data • System uses passwords or other security measures |
| <p>IV. COMP 2000</p> | <ul style="list-style-type: none"> • System constructed in modules • System has specialized storage structures <i>System embedded within a database, with the following features:</i> <ul style="list-style-type: none"> • historical data from previous surveys • System uses passwords or other security measures |

It should also be noted that all of the data editing systems permit linkages to some type of pre-existing data sets, either a sampling frame, historical data from previous administrations of the survey, and/or previously aggregated summary-level data. This feature is very important, because it permits comparisons to be made of the newly collected data with the either the sampling frame or earlier collected data. The comparison with the sampling frame allows non-response figures to be computed, and is especially useful if non-response rates require that imputations be generated. The ability to compare newly collected data with historical data or summary-level data from previous administrations also allows variations in response from administration to administration

to be evaluated, and provides the basis for the interpretation of new survey findings. National-office personnel are responsible for evaluating deviations from historical data sets and identifying reasons for these deviations.

IV. EDIT WRITING

Edit writing addresses those editing features and guidelines that help direct the editor or user in the usage of the existing editing system. Edits are usually written and designed to facilitate access to important information needed by data entry personnel. These edits, therefore, emphasize the importance of clarity in the data entry screen, and are designed to introduce important information to data entry personnel so that corrections can be made to data as needed. An example of this would be the flashing of an error message when an out-of-range value is obtained from a respondent. Another typical example would be the appearance of a historical variable that provides information on the respondent's past responses (usually from an earlier administration) for comparison purposes. The data entry person must evaluate the newly received data in the context of the historical information provided by the data entry system.

Another feature of edit writing allows the data editor/user to customize existing data editing features. Edit writing, therefore, is associated with the degree of flexibility in altering features present in the editing system, and the degree of latitude the editor can exercise in producing alterations to the existing editing program. Some of these alterations may serve to enhance the comfort level of the user, such as altering the display screen and any of its features (e.g., color, fonts, or highlighting) to facilitate legibility. Other alterations may provide the editor or data entry person with historical information so that an evaluation can be made about whether the data conforms to accepted criteria.

Table 5 provides information on the availability of edit writing features within the four BLS systems. All four editing systems allow historical and calculated variables to appear on the screen. Most of the editing systems ensure that variables are protected from change, while some of the editing systems allow visual display features to be customized as necessary. The ECI system also allows the user to exercise control over where the variables appear on the screen. All four editing systems also provide visual displays for errors in data; the most common display is an error signal that appears on the screen. However, the ARIES system uses pop-up windows, a feature which is likely to become increasingly popular at the BLS. Currently, many departments are considering altering their editing systems to accommodate a Windows environment that would permit the use of pop-up windows.

All four systems allow the user to test edits in an interactive or on-line mode (i.e., without paper output). In addition, the ARIES and SPAM editing systems allow you to test edits in batch mode, which generates paper output, if the user desires a hard copy of test edit results.

| TABLE 5 | |
|---|--|
| EDIT WRITING FEATURES OF BLS EDITING SYSTEMS | |
| I. SPAM System: OES | <ul style="list-style-type: none"> • Historical variables able to appear on the screen (may be viewed from Statistical Methods Division screen) • Calculated variables able to appear on the screen (may be viewed from Statistical Methods Division screen) • Variables are protected from change (some are able to be updated - other variables are not able to be updated) • Ability to change the color, fonts, or highlighting • Table format available • Errors able to be flagged through error signals on the screen |
| II. ECI Data Editing and Data Entry System | <ul style="list-style-type: none"> • Historical variables able to appear on the screen (prior quarterly data) • Calculated variables able to appear on the screen • Variables are protected from change • Ability to control where the variables appear on the screen • Ability to change the color, fonts, or highlighting • Calculations of auxiliary variables appear in messages • Errors able to be flagged through error signals on the screen |
| III. Aries Editing System: Current Employment Survey | <ul style="list-style-type: none"> • Historical variables able to appear on the screen • Calculated variables able to appear on the screen • Ability to change the color, fonts, or highlighting • Table format available <p><i>Errors able to be:</i></p> <ul style="list-style-type: none"> • flagged through error signals on the screen • viewed through pop-up windows • explained in text (e.g., "This value out-of-range.") |

| TABLE 5 | |
|---|---|
| EDIT WRITING FEATURES OF BLS EDITING SYSTEMS | |
| (Continued) | |
| | <ul style="list-style-type: none"> • Historical variables able to appear on the screen |

| | |
|---------------|---|
| IV. COMP 2000 | <ul style="list-style-type: none"> • Calculated variables able to appear on the screen • Variables are protected from change • Ability to control where the variables appear on the screen • Ability to change the color, fonts, or highlighting • Calculations of auxiliary variables appear in messages • Errors able to be flagged through error signals on the screen |
|---------------|---|

V. TYPE OF EDITS

The BLS editing systems are designed to apply a series of edits to all data collected either within the field by BLS personnel or obtained from respondents directly, via the respondent’s means. These edits are applied to all data to verify its accuracy, internal consistency, and consistency with historical records obtained through previous administrations.

Table 6 provides a summary of some of the types of edits that are conducted by the BLS editing systems. There is a large degree of overlap between BLS editing systems, because many of these types of edits are standard throughout the majority of editing systems. Edits range from qualitative to quantitative, and include numerous checks that ensure the accuracy of the data. Many consistency and balance checks are conducted to ensure that the data is internally consistent. These editing programs all have features designed to identify any discrepancies that may appear between two data pieces provided by a single respondent; when these errors appear, the editor must correct the discrepancy before the record will be accepted.

Similarly, these editing programs all have provisions to evaluate whether skip patterns present in the survey instrument have been handled properly, and ensure that all survey items have been properly administered and recorded. Additional edits include ensuring that incorrect figures, such as negative values are deleted; ascertaining whether linearity or non-linearity has been addressed, as needed; and that univariate and/or multivariate range checks have been performed as appropriate.

| TABLE 6: TYPES OF EDITS PERFORMED BY BLS EDITING SYSTEMS | |
|---|--|
| I. SPAM System: OES | <ul style="list-style-type: none"> • Linear edits (i.e., their graphs are straight lines) • Non-linear edits • Conditional edits • Ratio edits |

| | |
|---|--|
| | <ul style="list-style-type: none"> • Variables that can accept negative values • Quantitative edits • Univariate range checks • Multivariate, record-level checks for consistency • Consistency “balance checks” (e.g., parts sum to totals) • Other consistency checks (e.g., marital status vs. age) • Route checks or skip patterns • Check of generated values against external standards |
| <p>II. ECI Data Editing and Data Entry System</p> | <ul style="list-style-type: none"> • Linear edits (i.e., their graphs are straight lines) • Non-linear edits • Conditional edits • Ratio edits • Variables that can accept negative values • Quantitative edits • Qualitative edits • Univariate range checks • Multivariate, record-level checks for consistency • Consistency “balance checks” (e.g., parts sum to totals) • Other consistency checks (e.g., marital status vs. age) • Route checks or skip patterns • Check of generated values against external standards |

TABLE 6:
TYPES OF EDITS PERFORMED BY BLS EDITING SYSTEMS
(Continued)

| | |
|---|---|
| <p>III. Aries Editing System: Current Employment Survey</p> | <ul style="list-style-type: none"> • Linear edits (i.e., their graphs are straight lines) • Conditional edits • Ratio edits • Variables that can accept negative values (negative values are edited out) • Quantitative edits • Univariate range checks • Multivariate, record-level checks for consistency • Other consistency checks (e.g., marital status vs. age) • Route checks or skip patterns • Check of generated values against external standards |
| <p>IV. COMP 2000</p> | <ul style="list-style-type: none"> • Linear edits (i.e., their graphs are straight lines) • Non-linear edits • Conditional edits • Ratio edits • Variables that can accept negative values • Quantitative edits • Qualitative edits • Univariate range checks • Multivariate, record-level checks for consistency • Consistency “balance checks” (e.g., parts sum to totals) • Other consistency checks (e.g., marital status vs. Age) • Check of generated values against external standards |

Table 7 provides an overview of features of BLS editing systems that provide parameters regarding the nature of edits processed by these systems. All four systems accept logical operators. The ECI system has more than one level of edit priority, while the SPAM, ECI, and COMP 2000 systems can perform non-statistical cross-record checks. In addition, the foundation for the types of edits performed by each editing system can be found for each system. All four editing systems can perform statistical edits based upon historical data, cross record checks, and univariate outlier detection. Only the SPAM system can perform statistical edits based upon multivariate outlier detection.



TABLE 7**FEATURES OF BLS EDITING SYSTEMS**

| | |
|---|---|
| <p>I. SPAM System: OES</p> | <ul style="list-style-type: none"> • System can accept logical operators (e.g., ‘and,’ ‘or,’ ‘not’) • System can perform non-statistical cross record checks • States, regions etc. have customized error limits <p><i>Editing system can perform the following functions:</i></p> <ul style="list-style-type: none"> • check edits for redundancy • check that edits are not contradictory • generate implied edits • generate external records <p><i>System can perform statistical edits based upon:</i></p> <ul style="list-style-type: none"> • historical data • cross-record checks • univariate outlier detection • multivariate outlier detection |
| <p>II. ECI Data Editing and Data Entry System</p> | <ul style="list-style-type: none"> • System can accept logical operators (e.g., ‘and,’ ‘or,’ ‘not’) • System has more than one level of edit priority • System can perform non-statistical cross record checks <p><i>Editing system can perform the following functions:</i></p> <ul style="list-style-type: none"> • check edits for redundancy • check that edits are not contradictory • generate implied edits • generate external records <p><i>System can perform statistical edits based upon:</i></p> <ul style="list-style-type: none"> • historical data • cross-record checks • univariate outlier detection |

It should also be noted that all four editing systems are able to check edits for redundancy and ensure that edits are not contradictory. These important functions are used to treat problems with data records, such as identifying two records with the same unique identification code, or identifying records that are repeated. The latter check is designed to identify any contradictory pieces of information within any given record, and can sometimes be applied when cross-referencing across historical data.

In addition to the edits described above, all four BLS editing systems can generate both implied and external records. Implied edits are those that as a logical consequence must necessarily results because of the existence of at least two other edits. Thus, if one edit states that

a given value must fall between 3 and 8 and another value must fall between 10 and 12, the conjoint relationship of these two edits can yield an implied edit. Inspection of Table 7 indicates that BLS editing systems run the gamut through a comprehensive and extensive group of edits as necessitated by the given survey.

| TABLE 7 | |
|---|---|
| FEATURES OF BLS EDITING SYSTEMS | |
| <p>III. Aries Editing System: Current Employment Survey</p> | <ul style="list-style-type: none"> • 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 <p><i>Editing system can perform the following functions:</i></p> <ul style="list-style-type: none"> • check edits for redundancy • check that edits are not contradictory • generate implied edits • generate external records <p><i>System can perform statistical edits based upon:</i></p> <ul style="list-style-type: none"> • historical data • cross-record checks • univariate outlier detection |
| <p>IV. COMP 2000</p> | <ul style="list-style-type: none"> • System can accept logical operators (e.g., ‘and,’ ‘or,’ ‘not’) • System can perform non-statistical cross record checks • States, regions etc. have customized error limits <p><i>Editing system can perform the following functions:</i></p> <ul style="list-style-type: none"> • check edits for redundancy • check that edits are not contradictory • generate implied edits • generate external records <p><i>System can perform statistical edits based upon:</i></p> <ul style="list-style-type: none"> • historical data • cross-record checks • univariate outlier detection |

VI. DATA REVIEW AND CORRECTION

The data review and correction phase of data editing entails the review of newly collected data, often through the context of historical data collected through previous administration, and the correction of problems identified through this data review stage. For all four BLS editing software programs, the end-user is able to access the software program through menus. Menu-driven systems appear to make the data review and correction process more efficient and user-friendly.

Inspection of Table 8 indicates that all BLS editing systems allow manual review and the making of changes at all phases of the data collection process: before, during, and after data entry. It should also be noted that there are variations in the processes used by editing software regarding the data cleaning process. For ARIES, data cleaning occurs during data editing, whereas for COMP 2000 data cleaning occurs before and after data editing. Some of the editing programs require manual resolution of machine-generated signals, while others have batch capability and require manual resolution of errors.

| TABLE 8 | |
|---|---|
| STAGES IN BLS EDITING SYSTEMS | |
| I. SPAM System: OES | <p><i>Editing system allows manual review and changes:</i></p> <ul style="list-style-type: none"> • before data entry • during data entry • after data entry <p><i>Editing software:</i></p> <ul style="list-style-type: none"> • makes all corrections automatically in batch • requires manual fixes of machine-generated signals • has batch capability with manual resolution of errors |
| II. ECI Data Editing and Data Entry System | <p><i>Editing system allows manual review and changes:</i></p> <ul style="list-style-type: none"> • before data entry • during data entry • after data entry <p><i>Editing software:</i></p> <ul style="list-style-type: none"> • requires manual fixes of machine-generated signals |

| TABLE 8 | |
|--------------------------------------|--|
| STAGES IN BLS EDITING SYSTEMS | |
| (Continued) | |

| | |
|---|---|
| <p>III. Aries Editing System: Current Employment Survey</p> | <p><i>Editing system allows manual review and changes:</i></p> <ul style="list-style-type: none"> • before data entry • during data entry • after data entry <p><i>Editing software:</i></p> <ul style="list-style-type: none"> • requires substantial data cleaning during data editing process • requires manual fixes of machine-generated signals • has batch capability with manual resolution of errors |
| <p>IV. COMP 2000</p> | <p><i>Editing system allows manual review and changes:</i></p> <ul style="list-style-type: none"> • before data entry • during data entry • after data entry <p><i>Editing software:</i></p> <ul style="list-style-type: none"> • requires substantial data cleaning before it is applied • requires substantial data cleaning after it is applied • has batch capability with manual resolution of errors |

Table 9 provides additional information on operating characteristics of each of the BLS editing systems. All four editing systems permit data entry and data editing procedures to be conducted at the same time. All four systems permit high speed data entry onto a full screen; all but the ARIES system permit item code data entry, as well. Verification procedures in which data is entered twice can also be performed with all four editing systems; verification provides an additional safeguard during the data entry and editing process. The ECI and ARIES editing systems also permit data-entry statistics to be generated, which can be further reviewed to evaluate whether there have been irregularities within the data entry process.

Table 10 depicts complex data structures that can be used and manipulated by BLS editing systems. These more complicated data structures permit a higher level of editing to be conducted, by permitting the manipulation of complex and intricate data structures, such as complicated skip patterns, cross-record checks, subfiles, and hierarchical data. In addition, the SPAM and ECI editing systems allow respecification of both edits and data, which can be performed without additional manual changes to the editing software system. The combination of edits and complex data structure edit help ensure the accuracy of survey data.

TABLE 9

EDITING SYSTEMS: OPERATING CHARACTERISTICS

| | |
|---|---|
| <p>I. SPAM System: OES</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • data entry and data edit at the same time • data to be entered with correction (high speed) • full-screen data entry • item code data entry • verification (i.e., enter data twice) |
| <p>II. ECI Data Editing and Data Entry System</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • data entry and data edit at the same time • data to be entered with correction (high speed) • full-screen data entry • item code data entry • verification (i.e., enter data twice) • implied decimals to be typed (e.g., type 23 for 2.3) (used only for some fields) • data-entry statistics to be generated |
| <p>III. Aries Editing System: Current Employment Survey</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • 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 |
| <p>IV. COMP 2000</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • data entry and data edit at the same time • data to be entered with correction (high speed) • full-screen data entry • item code data entry • verification (i.e., enter data twice) |

TABLE 10

**COMPLEX DATA STRUCTURES USED AND
MANIPULATED BY BLS EDITING SYSTEMS**

| | |
|---|---|
| <p>I. SPAM System: OES</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • manipulation of subfiles • creation of subfiles • manipulation of cross-record checks (state boundary-limited; can also be used across establishments) • respecification of edits without other manual software changes |
| <p>II. ECI Data Editing and Data Entry System</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • 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 |
| <p>III. Aries Editing System: Current Employment Survey</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • manipulation of hierarchical data |
| <p>IV. COMP 2000</p> | <p><i>Editing system allows:</i></p> <ul style="list-style-type: none"> • manipulation of hierarchical data • manipulation of cross-record checks |

Once all data edits have been completed by the software editing systems, imputation procedures can be implemented as warranted, pending non-response rates and missing data. Typical imputation procedures include determining which variables to be changed, the automatic correction/imputation. These imputation procedures are selected based on problems with non-response and/or missing data, and include: hot-deck donor imputation, cold-deck donor imputation, administrative frame imputation, deterministic imputation, simultaneous multivariate imputation, and when necessary, several imputation options per variable. Currently, imputation procedures are only applied with the SPAM and COMP 2000 data sets.

VII. SUPPORT, UPDATES, AND TRAINING

The final component of the editing system describes the availability of support, training, and update information for users of the software editing systems. As may be seen in Table 11, the systems generally have some, if not all, of the following features available for obtaining additional assistance in using the editing software system and interpreting the output of editing and imputation procedures. These features include:

- On-line documentation
- User's guide
- Reference manual
- Manual of examples

These materials are available to assist data editors and end-users in implementing the editing system effectively.

| TABLE 11 SUPPORT, UPDATES, AND TRAINING AVAILABLE TO USERS | |
|---|--|
| I. SPAM System: OES | <i>Support, update, and training available to users:</i> <ul style="list-style-type: none"> • On-line documentation • User's guide • Reference manual • Manual of examples |
| II. ECI Data Editing and Data Entry System | <i>Support, update, and training available to users:</i> <ul style="list-style-type: none"> • On-line documentation • User's guide • Reference manual • Manual of examples |

| TABLE 11 SUPPORT, UPDATES, AND TRAINING AVAILABLE TO USERS | |
|---|--|
|---|--|

| | |
|--|---|
| III. Aries Editing System: Current Employment Survey | <i>Support, update, and training available to users:</i> <ul style="list-style-type: none">• On-line documentation• User's guide• Reference manual• Manual of examples |
| IV. COMP 2000 | <i>Support, update, and training available to users:</i> <ul style="list-style-type: none">• On-line documentation• User's guide• Reference manual• Manual of examples |

REFERENCES

- Baker, RP (1992). "New technology in survey research: Computer-assisted personal interviewing (CAPI). *Social Science Computer Review*, 10, 145-157.
- Bernard, C. (1989). "Survey data collection using laptop computers." Paris: INSEE (Report No. 01/C520).
- Cannell, C. F., Lawson, S. A., and Hausser, D. L. (1975). "A technique for evaluating interviewer performance." Ann Arbor, MI: Institute for Social Research.
- Catlin, G. and Ingram, S. (1988). "The effects of CATI on costs and data quality: A comparison of CATI and paper methods in centralized interviewing." Chapter 27 in Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L. and Waksberg, J. (eds.), *Telephone Survey Methodology*. New York: Wiley.
- Cotton, P. (October, 1988). "A comparison of software for editing survey and census data." A paper presented at the Statistics Canada Symposium, The Impact of High Technology on Survey Taking. (The topical division of this checklist and about half, perhaps more, of the items are taken from this report.)
- Couper, M. P. and Burt, G. (1994). "Interviewer attitudes toward computer-assisted personal interviewing (CAPI)." *Social Science Computer Review*, 12, 38-54.
- Couper, M. P., Hansen, S. E., and Sadosky, S. A. (1996). "Evaluating interviewer use of CAPI technology." Monograph submission.
- Couper, M. P., Sadosky, S. A., and Hansen, S. E. (1994). "Measuring interviewer behavior using CAPI." *Proceedings of the Survey Research Methods Section, American Statistical Association*, 845-850.
- "Data Editing in Federal Statistical Agencies." Statistical Policy Working Paper No. 18. Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget.
- de Bie, S. E., Stoop, I. A. L., and de Vries, K. L. M. (March, 1989). "CAI software, an evaluation of software for computer assisted interviewing." VOI, Association of Social Research Institutes, Data Collection Group, Amsterdam, the Netherlands.

- Dielman, L. and Couper, M. P. (1995). "Data quality in a CAPI survey: Keying errors." *Journal of Official Statistics*, 11(2), 141-146.
- Donmyer, J. E., Piotrowski, F. W., and Wolter, K. M. (1991). "Measurement error in continuing surveys of the grocery retail trade using electronic data collection methods." Chapter 18 in Biemer, P. P., Groves, R. M., Lyberb, L. E., Mathiowetz, N. A., and Sudman, S. (eds.), *Measurement Errors in Surveys*. New York: John Wiley & Sons, Inc.
- Frese, M. and Altmann, A. (1989). "The treatment of errors in learning and training" in Bainbridge, L. and Ruiz Quintanilla, S. A. (eds.), *Developing Skills With Information Technology*. Chichester: Wiley, 65-86.
- Greenberg, B. and Petkunas, T. (1987). "An evaluation of edit and imputation procedures used in the 1982 economic censuses in business division." 1982 Economic Censuses and Census of Government Evaluation Studies, U.S. Department of Commerce, 1987, 85-98.
- Groves, R. M. and Mathiowetz, N. A. (1984). "Computer assisted telephone interviewing: Effect on interviewers and respondents." *Public Opinion Quarterly*, 48, 356-369.
- Groves, R. M. and Nicholls, W. L. II (1986). "The status of computer-assisted telephone interviewing: Part II - data quality issues." *Journal of Official Statistics*, 2, 117-134.
- Kennedy, J.M., Lengacher, J.E., and Demerath, L. (1990). "Interviewer entry error in CATI interviews." Paper presented at the International Conference on Measurement Errors in Surveys.
- Lessler, J.T. and Kalsbeek, W. D. (1992). "Sources of survey error." Chapter Two in *Nonsampling Errors in Surveys*. New York: John Wiley & Sons, 13 -39.
- Little, R. and Smith, P. (1982). "Editing and imputation for quantitative survey data." *Journal of the American Statistical Association*, 82(397), 58-68.
- Martin, J., O'Muircheartaigh, C., and Curtice, J. (1993). "The use of CAPI for attitude surveys: An experimental comparison with traditional methods." *Journal of Official Statistics*, 9, 641-662.
- Nicholls II, W. L. and Groves, R. M. (1986). "The status of computer-assisted telephone interviewing: Part I-- Introduction and impact on cost and timeliness of survey data." *Journal of Official Statistics*, 2, 93-115.

- Norman, D.A. (1983). "Design principles for human-computer interfaces." *Proceedings of CHI '83: Human Factors in Computing Systems*. New York: ACM, 1-10.
- Norman, D. A. (1993). "Design rules based on analyses of human error." *Communications of the ACM*, 26: 254-258.
- Olsen, R. (1991). "Mode effects on data quality-- CAPI versus pencil and paper." Ohio State University, unpublished paper.
- Pierzchala, M. (1988). A review of the state and of the art in automated data editing and imputation. National Agricultural Statistics Service, United States Department of Agriculture.
- Rasmussen, J., Duncan, K., and Leplat, J., eds.(1987). *New Technology and Human Error*. Chichester: Wiley.
- Reason, J. (1990). *Human Error*. Cambridge: Cambridge University Press.
- Rogers, T. (1976). "Interviews by telephone and in person: Quality of responses and field performance." *Public Opinion Quarterly*, 40, 51-65.
- Rustemeyer, A. (1977). "Measuring interviewer performance in mock interviews." *Proceedings of the American Statistical Association, Social Statistics Section*, 341-346.
- Schneiderman, B. (1992). *Designing the user interface: Strategies for effective human-computer interaction*. (2nd ed.) Reading, MA: Addison-Wesley.
- Sebestik, J., Zelon, H., Dewitt, D., and O'Reilly, J. O. (1988). "Initial experiences with CAPI." *Proceedings of the Bureau of the Census Fourth Annual Research Conference*. Washington, DC: U.S. Bureau of the Census, 357-365.
- Tortora, R.D. (1985). "CATI in an agricultural statistical agency." *Journal of Official Statistics*, 1(3), 301-314.
- Weeks, M. F. (1992). "Computer-assisted survey information collection: A review of CASIC methods and their implications for survey operations." *Journal of Official Statistics*, 8, 445-465.
- Weinberg, Eve (1983). "Data collection: Planning and management." Chapter Nine in Peter H. Rossi et al. (eds.), *Handbook of Survey Research*. New York: Academic Press, 329-358.

Zapf, D., Brodbeck, F. C., Frese, M., Peters, H., and Pruemper, J. (1992). "Errors in working with office computers: A first validation of a taxonomy for observed errors in a field setting." *International Journal of Human-Computer Interaction*, 4, 311-339.