The Use of Paradata for Evaluating Interviewer Training and Performance October 2008

William Mockovak, [1] Randall Powers [2]

[1 & 2] Bureau of Labor Statistics, 2 Massachusetts Ave, N.E. Washington, DC 20212

Abstract

Audit files produced by CAI software present a record of the actions and entries made by interviewers as an interview is completed. These data provide objective feedback on the interviewing process, which can lead to changes in training or supervision. In this study, a software tool was developed that searched for and provided frequency counts of text that described selected interviewer actions. Survey production files were then analyzed to answer a series of questions including how interviewers chose to enter data (keyboard vs. mouse), when and how they navigated in a complex instrument, how interviewers handled edit messages when they were triggered, and how more complex sections of the instrument (e.g., tables) were handled. The advantages and disadvantages of using paradata in this context are discussed.

Key Words: Paradata, interviewer behavior, interviewer performance, interviewer training

1. Introduction

Some computer-assisted interviewing (CAI) software packages offer the option of providing a detailed history, or audit trail, which presents a record of the entries made and actions taken by interviewers as they complete an interview. This detailed information about the process of the interview had been referred to as *paradata* (Couper, 2000), although as Scheuren (2000) notes, paradata can be either macro or micro in nature, for example, response rates versus detailed interviewer actions.

The audit-trail data used in this study were captured in plain text files that displayed a chronological sequence of interviewer entries and actions for individual interviews. Specific interviewer actions were identified through the use of key words. As a simple example, if an interviewer paged up or paged down, the key words *Move Page Up*, or *Move Page Down* appeared in the audit trail. If the interviewer entered a don't know response to a question, the key words, *Don't know,* appeared by the question label. Similarly, entry fields (items/variables) in the instrument were identified in the audit-trail file, and entries made by an interviewer for that item were recorded, including occasions when a previous entry was deleted. However, not all interviewer actions were captured in the audit trail. For example, if an interviewer started entering an answer but deleted it and entered a new value before pressing the Enter key, this series of actions would not be captured.

Common uses of audit-trail data are to obtain timing data for individual items, sections, or an entire questionnaire; to review questionable entries for the purposes of data adjustment; or in rare cases to recreate an entire interview if a data file is corrupted. However, the primary objective of this research was to explore the value of audit trails as a supplementary measure of an interviewer's performance or behavior. This study focused on the first six months of field work after the conversion of a survey (Consumer Expenditure Quarterly Interview Survey) from paper to computer-assisted personal interviewing (CAPI). The Quarterly Interview survey is an ongoing, panel survey that collects detailed information about monthly out-of-pocket expenditures such as housing, apparel, transportation, health care, insurance, and entertainment. It was hoped that measures obtained from the audit trails could provide useful feedback about the effectiveness of training, the quality of the interview process, decisions made about instrument design, and the general usability of the instrument. Additionally, it was hoped that the paradata could be used to improve the efficiency and quality of field work.

2. Method

Although it's possible for a person to read an audit trail after some training, to be maximally useful and to generate a statistical summary, the cryptic information in the file first needs to be summarized. To accomplish this, we developed a tool that enables the user to search for specific strings of text in the audit-trail files so that the frequency of that text string's occurrence could be determined. Using this approach, 26 different interviewer actions were analyzed. In addition, timing data for the interview and individual sections were obtained using a different method. In the interest of brevity, only selected results will be presented in this paper. Since a paper-based survey had just been converted to CAPI, the period of time after implementation was of immediate interest, so data from five of the first six months after the survey converted from paper to CAPI were analyzed. The analysis included all interviewers, including replacement interviewers. Although turnover among CE interviewers obviously occurs, it was not viewed as a major problem in this study. Interviews were conducted on standard size laptop computers. Depending on the interview (there were five separate interviews), an average CAPI interview lasted between 50 and 60 minutes. The number of cases analyzed over a 6-month period ranged from 5,294 in April to 5,389 in September for a total of 22,427.

2.1 Analytical Objectives

Mode of Data Entry. During the design of the CAPI instrument, a debate occurred over which mode of data entry interviewers were most likely to use: the keyboard or the mouse (a trackpad and "pencil-eraser" mouse were both available on the laptop). The argument for keyboard entry was that interviewers occasionally had to conduct interviews standing up (or outside) and, even when inside, suitable surface areas could be limited. Although the instrument was designed to allow both modes of entry, the classroom training focused on keyboard entry. Since some field supervisors questioned this strategy, a key objective was to determine which mode of entry interviewers used, so that future training could be adjusted, if necessary.

Navigation. As part of the conversion from paper to CAPI, a variety of assumptions were made about the design of the instrument and data entry with the objective of enhancing usability. Since a typical Quarterly interview with a paper questionnaire lasted over an hour, a major concern in designing the CAPI interview was not to add significantly to the length of the interview. Therefore, the instrument was designed with a variety of navigation options and design features to give the interviewers as much flexibility as possible to try to recreate the freedom of movement found on the paper form, since interviewers specifically requested this functionality. For example, interviewers had asked for the ability to freely jump around among non-dependent sections of the instrument so that respondent burden could be reduced as they completed the 25-module instrument. In addition, because the CAI instrument used a large number of tables for collecting information, special data entry and navigation tools were developed specifically for those tables. These tools allowed time-saving actions such as automatically copying a cell of data in the table from one row to another row directly underneath, and function keys that would take the interviewer from any point in the table to the end of the table with one key press.

Because the CAPI instrument had 25 separate sections or modules, there was a debate about the best approach for enabling interviewers to skip among them. Two approaches were made available to the interviewer. One approach displayed tabs across the top of the screen (one tab for each section), so that clicking the tab would take the interviewer to that section. However, because there were so many tabs, not all of the tabs could be displayed at once. Instead, to see the tabs for later sections, interviewers either had to proceed past a certain point in the interview or they had to click a right-scroll button at the top of the screen. As an alternative, a drop-down menu existed that was triggered by pressing a function key (F4). When an interviewer pressed F4, the drop-down menu displayed all the sections in the instrument. The interviewer could then either use the arrow keys (and the Enter key) or a mouse/trackpad to select a section to jump to.

Edit Messages. One of the stated advantages of CAPI is that data can be corrected at the source, but questions about implementation and respondent burden have always remained concerning the optimal use of edit messages. For example, since edit messages clearly interrupt the flow of the interview, how often should they be used, how sensitive should they be (for value checks), and how do interviewers and respondents handle the two basic types: within an item (for example, a range check) or between items (a previously entered value conflicts with a current entry)? Effective wording of the edit messages is another issue. Do edit messages perform an adequate job of explaining what a problem is and how to correct it?

Other Quality Measures. The notes that interviewers enter to explain entries are an indirect measure of quality. These notes are used by program office staff to adjust the data when necessary, so their frequency of use is important, and they may also reveal problem areas in the instrument. Similarly, the frequency of *don't know* and *refused* entries are another indirect indicator of data quality, since they reflect the amount of missing data.

3. RESULTS

When given a choice, how do interviewers enter the data? As noted previously, classroom training focused on keyboard entry under the assumption that positioning and clicking the mouse (only "built-in mice" were provided) would be more time consuming and, therefore, more likely to add to respondent burden by lengthening the interview. So, did interviewers tend to use the keyboard or the mouse/trackpad when completing the instrument?

They used the keyboard for 92 percent of their entries, with only 8 percent being made using a mouse. (The range of entry fields completed varied from a low of 441 in July to a high of 502 in April.) It should be noted that the number of entry fields is not an accurate count of the number of separate items completed, because some fields might have been counted more than once. For example, if an interviewer backed up and came forward using one simple navigation approach, the items that were completed and then forwarded over would be counted again.

Navigation: How often, and which tool, do interviewers use to jump from one section to another? Navigation is of special interest because working with a complex instrument is probably the hardest CAPI skill to master, besides requiring a lot of classroom training time to cover. Moreover, past experience has shown that interviewers, even those with a lot of computer-assisted interviewing experience, often do not fully utilize all the navigational tools provided to them. Or, if they use them, they use them inefficiently. As noted previously, there were two basic ways to jump among sections. First, the interviewer could use the keyboard by pressing the F4 key, which results in a drop-down menu that lists all the sections. The interviewer could then use the mouse/trackpad to click on the appropriate section or use the arrow keys to highlight a section before pressing the Enter key. The other alternative for jumping to a section was to use the mouse/trackpad to click on tabs that appeared at the top of each screen. These tabs were labelled and represented each of the 25 sections of the instrument.

Figure 1 shows the frequency of use of the menu (F4) or the tabs for navigation. It is clear that the tabs, which required use of the mouse/trackpad, were the preferred method of jumping among sections, since they were used about six times more frequently than the drop-down menu. It is also worth noting that interviewers did jump among sections; slightly more than 2.5 times per interview, on average. It is possible this figure could be inflated slightly by jumps to the appointment section of the instrument. However, this seems unlikely since interviewers have a special function key (F10) that takes them directly to the appointment section, and they receive special training on its use. The F10 function key was used about 1.5 times per case (not shown in graph), although in about 10 percent of the cases it was used four or more times. This frequency of usage probably reflects how often a case was opened (the average was 2.6 times). The number of times a case was opened and use of the F10 key could be indirect measures of the difficulty of an interview since cases with cooperative respondents might be completed more readily on the first attempt.
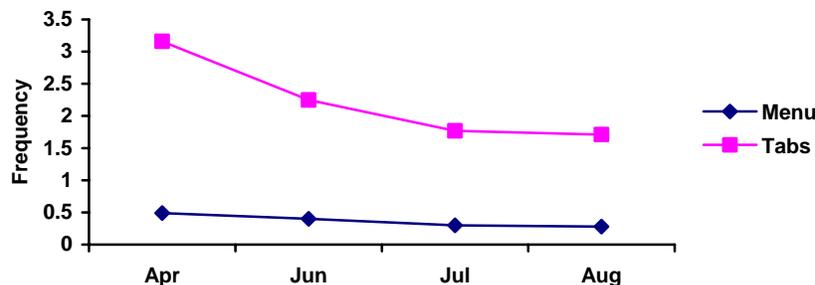
Figure 1. Frequency of Use of a Drop-down Menu (F4) or Tabs for Jumping Among Sections (Completed Interviews)

Jumping among sections was more frequent in the beginning months of CAPI use. Jumping appears to have started leveling off in July, the fourth month of production interviewing. Cases in which the interviewer started and finished the interview following the main logical path through the interview without any jumping among sections accounted for about 54 percent of the total cases. (Data for completed cases were not available for September.)

Use of additional ways to navigate – the *End* and *Home* keys

An explanation of the behavior of the *End* and *Home* keys requires a working knowledge of the concept of a *main path* in a Blaise instrument[1] and what happens when the interviewer leaves the *main path*, which happens any time a jump to another section occurs (either a forward or backward jump). Because it is a relatively involved concept, a simple explanation is that the *main path* is the unbroken sequence of questions from the very beginning to the end of the interview that results when questions are asked and answers are entered following the pre-specified logic in the instrument.

There are times when the *End* key is especially useful. For example, in any interview after the first interview, the interviewer will encounter a table of household demographic information with entries that need to be updated or verified. If no changes are required in this table, pressing *End* will move the cursor to the last entry of the table. Another way to get past this table is to repeatedly press a right or down arrow key, press tab or Enter, or use the mouse, but these actions are much slower that a single press of the *End* key. Another possible use of the *End* key occurs when interviewers back up to previous questions to check or to correct a previous entry. After doing so, pressing *End* will take them back to the next unanswered question (typically the question they were on when they decided to back up).

The *Home* key has less obvious value. If you press *Home* when you are on the *main path*, you will be taken to the first item in the instrument. If you make a jump off the main path to another section and press *Home* in that section, *Home* will take you to the first item in that section. However, the *Home* key could be used in this latter situation to orient the interviewer to that section and to the questions that had already been completed.

Neither the *End* key nor the *Home* key was used very often, although the *End* key was used more frequently. On average, the *End* key was used slightly more than once per case (1.04) in the first month of CAPI implementation. Its use then declined during the first two months and reached a plateau of about 0.7 times per case, whereas use of the *Home* key was highly variable (ranged from 0.06 to 0.85). Since the *End* key is a potential time saver, it's not clear why its use dropped off. It's possible that some interviewers were unaware of it or possibly forgot about using it in later months.

Use of the "Repeat" Key. Another tool (not a navigation tool) that interviewers specifically requested was the F12 or *Repeat* key. When used, this key copies the information from one cell in a table to the cell immediately beneath it. In the section that asks about utility expenses, the name of a utility or cable company usually does not change from month to month; therefore, this tool can save time retyping the information. The utility of this tool was evident. It was used about 14 times per case in April and over 17 times per case in August, when its use started to plateau.

How often are edit messages triggered and how do the interviewers handle them?

Blaise uses two basic types of edit messages: soft and hard. If a hard edit is used, the interviewer must change the answer (entry) that caused the edit message to appear. A soft edit message gives the interviewer three options:

1. *Suppress* – if selected, the value (answer) is saved, and the next question in logical sequence appears. By clicking *Suppress* the interviewer is essentially saying "the entry is fine … continue to the next question."
2. *Close* – if selected (or if the Escape key is pressed), the edit message disappears and the cursor remains on the entry field where the value (answer) had just been entered. The interviewer can now enter a different answer. However, if he/she enters the same value, the edit message will appear again.
3. *Goto* (or jump) – if selected, the interviewer will be taken to the question highlighted in the *Questions Involved* section of the edit dialogue box. The "goto" or *jump* option is valuable for edits that involve more than one item in the instrument. When only one item is involved in an edit, *close* and *goto* have the same outcome.

To estimate the number of times that edit messages appeared, counts were obtained of the number of times that messages were suppressed, closed, or a jump resulted. It's possible that this approach might result in a slightly inflated count if interviewers close an error message but then try to enter the same value. However, since this action was covered in training, and it would result in a delay in the interview, it's not assumed to have occurred very often. Table 1 shows the frequencies with which different actions were taken in response to edit messages. The sum of these actions is an estimate of the average number of times edit messages appeared during an interview.

[1] Blaise: A computer-assisted survey processing and interviewing system developed by Statistics Netherland.

Table 1: Average Number of Edit Actions in an Interview (all cases)

Action	April	June	July	August	September	Average
Suppressed	2.15	2.10	1.94	1.99	2.10	2.08
Goto (jump)	2.34	1.62	1.38	1.39	1.26	1.64
Close/Escape	0.99	0.66	0.52	0.54	0.47	0.66
Total	5.48	4.38	3.84	3.92	3.83	4.38

*Audit files were not available for May.

The total number of edits appearing was highest in the initial month of the conversion, April (5.5). By the fourth month (July), the number of edits had leveled off to about 3.8 per case, which suggests that the interviewers learned how to avoid some of the edits. The frequency of use of "suppress" by interviewers remained relatively constant throughout the entire time period (although the number of edit messages decreased), and the use of "goto" and "close" showed an initial drop after April, then a leveling out in the fourth month of implementation.

Did interviewers' reactions to the edit messages change with experience? The next chart shows that with increasing experience interviewers used the *suppress* option relatively more often when an edit message appeared. The use of *suppress* went from about 40 percent in the first month (April) to about 55 percent in the sixth month. On the other hand, the use of *goto* or jump dropped from about 42 percent to 33 percent over the same time period. The use of *close* also dropped. This pattern suggests that interviewers paid less attention (i.e., took less direct action) when edit messages appeared as they gained experience with the instrument.
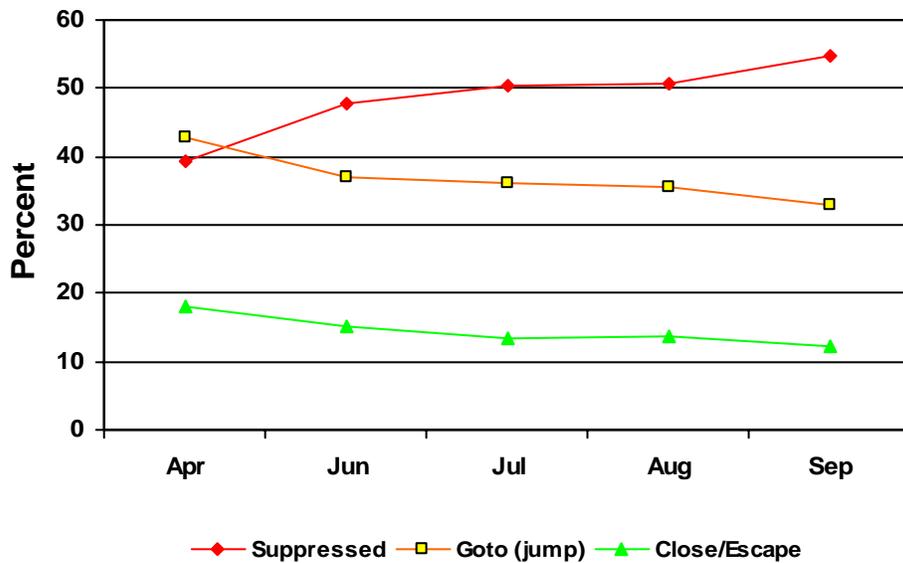
Figure 2. Actions Selected by Interviewers in Responding to Edit Messages

Slightly more than half of the cases had no suppressed edits (50.6 percent). On the other hand, one case had 52 edit suppressions. These results suggest that edit messages can pose serious usability problems for some interviewers, so they need to be handled carefully in training.

How often do interviewers enter notes? An interviewer could enter an item-specific note by pressing the F7 key. According to the analysis of the audit files, notes are used an average of less than once per interview (0.6-0.7).

How often are *don't know* and *refused* used?
The amount of missing data (either blank items or items with a don't know or refused entry) can be viewed as a basic measure of data quality. As the following graph shows, a *don't know* answer is entered between 6 and 8 times an interview (Average = 7.2). A *refused* answer occurs much less frequently, between 2 and 3 times (Average = 2.7); however, the standard deviations for these measures are both relatively large (sd_{dk} = 18.5; sd_{ref} = 19.1). As shown in the next graph, the use of *don't know* decreased during the first four months following implementation of CAPI, before

rising slightly. It's interesting to conjecture why this decrease occurred. It's possible that when interviewers were less experienced they encountered situations where entering a don't know answer was the easiest way for them to advance in the instrument. With increased experience, these situations may have decreased in number.
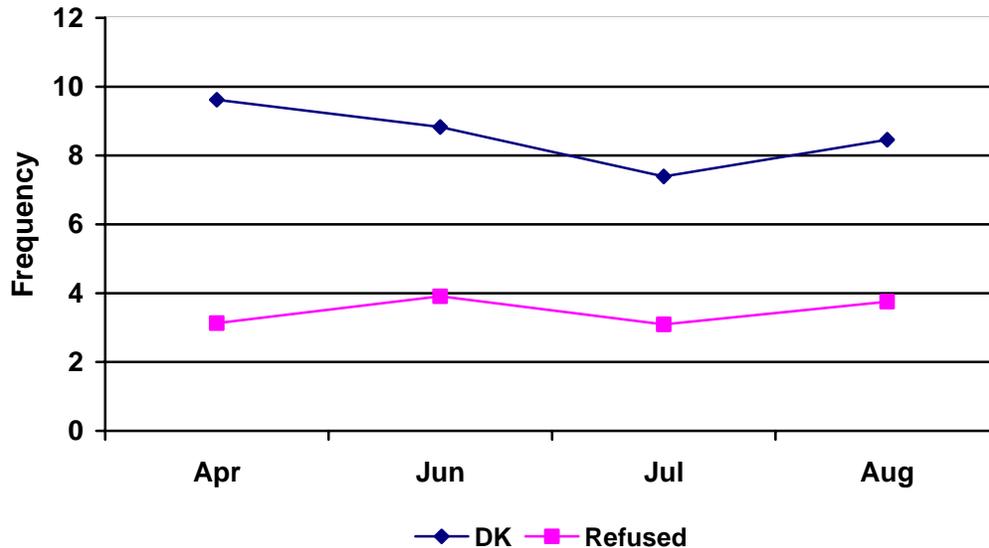

F**igure 3**. Use of Don't Know & Refused (Completes)

4. DISCUSSION

This research used paradata to answer a variety of questions about how interviewers interacted with a complex CAPI instrument and to describe how that behavior changed over time. The inherent appeal of paradata is that it is produced automatically, it's abundant, and it provides a detailed, objective record of the process of the interview. However, there are drawbacks. These include its shear volume and the fact that it must first be summarized before it can be used to answer either general or specific questions. This drawback has been noted before (Safir et al., 2001). For example, the CE Quarterly Interview survey generates about 5,000 audit trails each month, totaling about one gigabyte in their native format. We developed a software tool to summarize these data, but similar tools are available for other computer-assisted interviewing packages (Bumpstead, 2001).[2]

Although paradata might be useful in investigating general behaviors and possible inefficiencies, they are only a first step in the process of understanding why these behaviors occur. For example, as interviewers gained experience with the CAPI instrument when edit messages appeared did they increasingly press the suppress key (to make the message disappear) because the edit message was too sensitive, the message was viewed as an unwelcome interruption in an already long interview, or did they simply fail to understand the message? The only way to answer these questions is to talk with some interviewers and ask them.

The current research was exploratory. One practical objective was to assess the usefulness of paradata versus the amount of effort it took to analyze the data. In the process, we encountered operational problems with both our in-house analytical tool and the audit files themselves. Nonetheless, we were able to analyze an ample amount of data. So what did we learn?

As a practical result, the analysis provided us with information that was used to design an interviewer refresher training session held about nine months after CAPI was implemented. This training session addressed specific performance issues that had been noted in the paradata. In addition, the analysis allayed concerns about design decisions that had been made in the instrument and which affected its usability. The following is a brief summary of major conclusions, which should be considered in terms of the context of the instrument that was evaluated (a large, 25-module instrument requiring an average of 50-60 minutes to complete depending on the interview):

---

[2] For an analysis tool, see WesAudit for Blaise: http://www.westat.com/statistical_software/blaise/blaise_services.cfm

4.1 Findings on the Use of Navigational Features and "Special Purpose" Tools

- *Interviewers used the jump capability.* Interviewers had requested the ability to move freely among the non-dependent sections of the instrument, and they used that capability in slightly less than half of the cases (46 percent). Interestingly, with increased experience less jumping among sections occurred. Moreover, a majority of cases (54 percent) were completed without any jumping at all, so it appears that interviewers used the ability to jump judiciously.

- *Tabs were preferred for moving among sections.* Interviewers preferred using tabs to a drop-down menu when navigating among sections, even though use of the tabs required interviewers to use a trackpad or "pencil-eraser" mouse.

- *Specifically requested tools were used frequently.* During the iterative development of the instrument, interviewers had requested the use of a special key (repeat) to copy information from one cell of a table to a cell directly underneath (an obvious time saver). Use of this tool started out high and increased over time, so it obviously met a clear need.

- *Interviewers may forget to use special tools.* Special tools to aid in the completion and navigation of complex sections of the instrument (for example, tables) were used. However, the data suggested that the use of one special tool (*End* key) actually decreased over time. Since the occasions when this tool could be used were limited, it's possible that interviewers either forgot about its availability, figured out an equally effective way to accomplish the same action, or simply did not need to use it as much since less jumping occurred with more experience.

4.2 Other Findings

- *The keyboard was preferred for data entry.* The keyboard was used almost exclusively to enter data. This was not a surprise since its use was emphasized in training, but some field supervisors had strongly argued that the trackpad/mouse would ultimately prove to be more popular. However, it's worth noting that interviewers tended to use the trackpad or "pencil-eraser" mouse for jumping. This suggests that given a choice interviewers will use the tool that is easiest for them to use.

- *Edits may suffer from poor usability.* In preliminary work, some interviewers complained that edits were appearing too frequently and were too sensitive. Other concerns were expressed about their interpretability during the demands of an interview. The results suggest that the number of edits did not appear at an unmanageable rate. Although the analysis did not address the issue of an edit's sensitivity (for example, being triggered by values that may be too low or high), as interviewers gained experience they increasingly used "suppress" to make edit messages disappear, which may indicate that the edit messages are not as effective as they could be. For this reason, edit messages demand further study to determine if their usability and, hence, value could be improved.

- *The use of "don't know" responses decreased over time.* Although the overall frequency was fairly low to begin with, the use of *don't know* entries decreased during the first four months following implementation of CAPI, before rising slightly. This suggests that familiarity with the instrument leads to fewer *don't know* entries, but the effect is not large, only about two entries per case.

Because this was preliminary research, there were some key questions that were not answered. A partial list follows:

- Do edits tend to appear more frequently for certain items (or in certain sections)?
- How much time do edits add to the interview?
- What percentage of the time does an edit message result in a changed answer?
- Which types of interviewers (experienced, highest response rates, etc.) tend to use the more difficult navigational tools? Is use of these tools correlated with the length of the interview?
- How often do interviewers back up and change an answer (or back up and not change an answer)?
- Given the significant amount of time devoted to creating and updating help screens, how often do interviewers use them, and is their use indicative of a possible problem in a section?
- Timing data can also be obtained from audit-trail files. Are differences in completion times for different sections of the instrument (or the entire instrument) correlated with differences in paradata?
- Did interviewer turnover affect any of our major conclusions?

As noted previously, we encountered operational problems with both our in-house analytical tool and the audit files themselves, so analyzing the audit trails was not an easy process. One of our major conclusions was that a simpler, less time consuming approach was necessary if audit trails were to be used on a more routine basis for survey management decisions. Therefore, as a follow-up to our work, additional work was done at BLS by Chan et al. (2004) to make the audit trails easier to use and to answer some of the questions that we had been unable to address. Other organizations have also followed this route and have developed systems that can be more easily used to access and analyze the information in audit trail files, in one case to provide more objective measures of interviewer performance.[3]

In summary, although paradata have proven useful in the current context, to be of maximum benefit they must be produced more routinely and simply so that their possible impact on measures of data quality can be more easily explored.

References

Bumpstead, R. (2001). A Practical Application of Audit Trails. Paper presented at the annual Blaise Users Conference. WEB LINK www.blaiseusers.org/ibucpdfs/2001/Bumpstead--IBUC_paper.pdf (April 17, 2008).

Chan, W.C., Ho, P., Kang, S., Smith, L., and Tan, L. (2004) CE Interview Audit Trails: Report on development of a prototype for data storage and analysis. BLS internal report.

Couper, M. P. (2000) Usability evaluation of computer-assisted survey instruments, Vol. 18, No. 4, pp. 384-396.

Safir, A., Black, T., Steinbach, R. (2001) Using paradata to examine the effects of interviewer characteristics on survey response and data quality. Paper presented at the annual meeting of the American Statistical Association, August 5-9.

Scheuren, F. (2000) Macro and micro paradata for survey assessment. http://www.unece.org/stats/documents/2000/11/metis/crp.10.e.pdf

---

[3] POINT System at Statistics Canada, personal communication with Michael Egan.