

USING TRACE FILES TO EVALUATE THE NATIONAL HEALTH INTERVIEW SURVEY CAPI INSTRUMENT

Mick P. Couper (University of Michigan & Joint Program in Survey Methodology),
John Horm (National Center for Health Statistics), Jay Schlegel (University of Michigan)
Mick Couper, P.O. Box 1248, Ann Arbor, MI 48106

Key Words: Computer assisted interviewing, CAPI, keystroke files, trace files

Introduction

The advent of computer assisted interviewing (CAI) potentially requires greater skill on the part of the interviewer in interacting with the computer. Particularly with the DOS-based CAI systems still dominating the field, completing the instrument may not be as intuitive a process as was the case in paper and pencil surveys. This may be especially true when deviations from the standard forward flow through the instrument are required. However, the automation of the interviewing process not only offers greater complexity and potential difficulty for the interviewer, but at the same time offers a means of evaluating how successfully interviewers carry out the task of interviewing with a computer.

Trace files or keystroke files are automatic byproducts of most CAI software systems. While they are generally used for diagnostics and debugging instruments during testing and development, trace files can also be used to evaluate interviewer use of the CAI system (see Couper, Hansen, and Sadosky, 1997). In this paper, we explore the use of trace files in examining what functions interviewers use (or do not use) and when and how they use these functions. In addition, trace files can be used to identify problems with the instrument. Trace files are roughly analogous to other micro-level data on the interviewing process, such as may be captured through monitoring or taping and behavior coding. While the latter methods are rich data sources for understanding interviewer, respondent or instrument difficulties, they are expensive and time-consuming to collect. Trace files are virtually costless to collect, and are available almost instantaneously, making them very useful for evaluation of the instrument during pretesting, or as a source of relatively quick feedback on the effectiveness of interviewer training.

Trace or keystroke files vary in the level of detail they provide. In transaction-based CAI systems (e.g., Surveycraft), every key that is pressed by the interviewer is recorded in the keystroke files. On the other hand, in execution-based systems (like CASES), only those entries that are executed by the system are captured in the trace files. Trace files usually include question identifiers, permitting question-level analysis. Furthermore, some

systems permit the capture of timing information associated with each question (e.g., time in 1/100 seconds), permitting detailed and accurate analysis of how long each question may take. However, such rich detail may come at a price — the more information that is captured, the larger the resulting files. This may slow down the system during data collection, and increase the costs of transmission and storage. One of the challenges in the analysis of trace files is that their sheer number has the ability to rapidly overwhelm the analyst. Thus, finding ways to summarize the files and extract the pertinent information has been a key goal from the outset (see Couper, Hansen, and Sadosky, 1997).

Furthermore, regardless of the amount of information captured, trace files can only reveal *what* happened at a particular moment, not *why*. The data require interpretation to understand what the interviewer and respondent were doing at the time, and the reasons behind a certain sequence of events. Therefore, we believe that trace file analysis is most usefully done in combination with some other method (e.g., field observation, laboratory testing, behavior coding or interviewer debriefing) that will help us to understand what is going on in the interview and why. In our case, we are combining the trace file analysis with usability testing (see Hansen, Fuchs, and Couper, 1997).

In this paper we examine one particular source of trace files, namely the National Health Interview Survey (NHIS), programmed in CASES version 4.2. We first describe the survey and trace file data in more detail, then explore the analysis of the trace files at different levels (e.g., interviewer, case, question) to address different issues. We present a variety of examples to illustrate the kinds of information that can be gleaned from trace files and the utility to which they can be put.

Data Sources

The data we examine are from the CAPI version of the National Health Interview Survey (NHIS). The NHIS is an ongoing survey of health-related issues in the United States, and has been conducted continuously since 1957. The survey has undergone a period of redesign over the past few years as it converted from paper and pencil interviewing (PAPI) to computer assisted personal interviewing (CAPI). The redesign also involved extensive alterations to the questionnaire, and changes to

the sample design. Phase I of the redesign, conducted during the first six months of 1996, involved a small number of interviewers (16) using the CAPI instrument. In Phase II, during the second six months of 1996, all NHIS interviewers conducted about half their sample cases using CAPI, while the balance was done using PAPI. Finally, in Phase III, beginning in January 1997, the entire NHIS sample was switched to CAPI, with no more paper interviews being conducted.

The trace file data we report are from Phase II of the NHIS. During this six-month period, several versions of the instrument were released (approximately one every six weeks). We have trace files from 16,725 completed interviews, containing 3,045,611 questions. This represents an average of 182 completed questions per interview. A total of 370 different interviewers conducted CAPI interviews in this period, with an average of 45 cases per interviewer.

Analyses

We first examine the use of various CAPI functions, at both the case level and interviewer level. Aside from cursor key movement and the [Enter] key for executing an entry, all interviewer actions make use of functions keys. Table 1 shows the use of various functions in the NHIS instrument. Using this information, we can tell that virtually all interviewers know how to use the [F1] key to back up in the instrument. This is also a frequently used function, with an average of over 13 occurrences per interview. However, [F2] is much less frequently used, probably because one could move forward again (after backing up) simply by pressing [Enter] at each answered item until the next unanswered item is reached. For lengthy backup episodes, we would expect interviewers to use [F3] to jump to the next unanswered question. This function is used relatively infrequently, probably because the distance of most backups is not sufficiently large to necessitate its use. On the other hand, it appears that interviewers are familiar with this function, with 87% having used it at least once.

Interviewers were trained to enter question-specific notes to clarify answers that did not fit into the response options, or otherwise present additional information. Fully one-third of the interviewers appear never to have used this function, and in general this is used rarely. This could suggest a lack of knowledge of the purpose of this function on the part of the interviewer; however, we suspect it is more likely that interviewers feel such comments are unlikely to be read by supervisory or survey staff, and entry of comments only serves to delay the progress of the interview. Admittedly this is speculation, and this issue deserves further exploration. However, we have seen large variation in the use of

question-specific notes across surveys, suggesting that differences in interviewer training and possibly also in instrument content may account for variation in the use of such notes.

Very few interviewers use [Sh-F10] to display the list of function keys. This makes sense for two reasons. First, a paper function key template was provided to interviewers to attach to the laptop. Second, if the interviewer needs help remembering function key mappings, it is unlikely that they would remember to press [Sh-F10] in order to obtain the needed information. The presence of the template suggests that simply not knowing what the functions are is not sufficient reason for their lack of use. More likely explanations are that interviewers do not have a need for these functions, or are uncertain about the exact operation or consequences of a particular function, and therefore avoid its use.

We can also examine the lack of use of those functions we would expect interviewers to use. For example, in a roster verification screen (illustrated in Figure 1), the names of the first three household members are displayed. Interviewers are required to read all names, and, for households with four or more persons, are expected to press [Sh-F6] to activate the display of names (the lower window in Figure 1), then [PgDn] to see the remaining names. In our usability observations (see Hansen, Fuchs, and Couper, 1997) we have seen that few interviewers seem to be able to do this. They employ a variety of different strategies to overcome this problem, including remembering the names, writing the names on a piece of paper, enlisting the respondent's help, or not reading the names at all. From the trace files we see that [Sh-F6] was used in only 1.7% of all interviews (see Table 1). However, we only expect this function to be used in households with four or more persons. Among such households, comprising 19.1% of the interviewed cases, [Sh-F6] is used once or more in only 7.2% of cases. Put another way, in 92.8% of the interviews in which we expect this function to be used, it is not used. This suggests an operation that is unnecessarily complex, or insufficient interviewer training on the use of this function.

Furthermore, interviewer-level use of this function has declined over time (as measured by contrasting the four versions of the instrument). In version 2.1, 28.9% of interviewers had used [Sh-F6] at all, and this dropped to 24.0% in version 2.2, 10.4% in version 2.3 and 12.4% in version 2.4. Similar declines are seen in interviewer use of other functions, e.g., [F3] to jump forward (from 83% down to 55%), [F6] to change entry (from 32% to 18%), [F4] for the jump table (from 42% to 22%).

Turning to the use of item-specific help, there are a total of 384 different help screens in the Phase II NHIS

instrument, accessed by pressing [H], displayed on those screens that have help text available. In the first 6 months of CAPI for the NHIS, only 203 of these screens had ever been accessed, and only 114 accessed more than once. Only 9.4% of all cases have any help access. This parallels findings by Baker (1992), Couper, Sadosky, and Hansen (1994) and Sperry et al. (1996) that use of help is rare across systems and surveys. Examining these data further, we find that five help screens account for 67% of all help access. Two of these are in the case management portion of the instrument (call history, accounting for 27% of all help access; and noninterview codes, 11% of all help). The most frequently used help item in the interview itself (14% of help access) was one on state of birth, where [H] brought up a list of 2-digit state abbreviations. A navigation grid (accessed using [Sh-F3]) was next most frequently used (12%), followed by text explaining why the survey needed to records respondent's social security numbers (5% of help access). These give an indication of the types of difficulties encountered by interviewers as they conduct the survey. In fact, the navigation grid function ([Sh-F3]) appeared on the function key template, but was not activated in the early part of Phase II. Not surprisingly, we see dramatic declines in attempted use of this function, from 4.6 occurrences per interview in version 2.1, to 2.6 occurrences in version 2.2, 0.2 in version 2.3 and none in version 2.4. Again, this suggest a possible need for a navigation grid or overview of progress, one that was not provided in the instrument at the time (see also Sperry et al., 1996).

There are several reports that backing up to change previous answers presents particular challenges for interviewers (e.g., Couper and Burt, 1994; Sperry et al., 1996). We separately analyzed all backup episodes in the Phase II trace files to explore a variety of questions: (1) do interviewers back up one question at a time or use jump back, (2) what are the primary targets of backup actions, and (3) what do interviewers do when they backup to an item?

We find that item-by-item backup (using [F1]) is common, occurring an average of 13.6 times per interview, and occurring in 74% of all interviews (see Table 1). The two jump back functions ([F4] and [F5]) are more rare, with an average of only 0.08 per interview for [F4] and 0.14 for [F5]. In each case, these functions are used in less than 3% of all interviews, and by only about half the interviewers. This may even overestimate the use of backup episodes, as in the NHIS we cannot distinguish between functions generated automatically by the system and those generated by the interviewer pressing a function key.

Exploring single-item backups (using [F1]) in more

detail, we find use of this function declines slightly over time (from an average of 15.3 per interview in version 2.1 to 11.8 in version 2.4). In addition, we find that 19.1% of all backups are within the same screen. This suggests that interviewers are using the backup function rather than cursor keys to move around items in multi-item screens. Finally, we find that the top ten targets of backup episodes account for 16.6% of all backups. We examined some of these in more detail to explore the reasons why backups are taking place.

The most frequent target of backups is a mark all that apply screen on types of health insurance coverage. The design of this screen was changed during Phase II, requiring an interviewer to place an 'X' next to the selected items, in versions 2.3 and 2.4. In the earlier versions (and in all other multiple response items), interviewers were required to enter the number corresponding to each selected item. Of the 3,375 backups to this screen, 45% are movement across items within the screen, and, of these, in fully 61% of the instances, the interviewer is changing a number to an 'X'. This clearly suggests that interviewers are attempting to use this screen as they would any other mark all that apply item in the instrument, requiring additional effort to identify and correct the problem. Most of the remaining backups to the health insurance screen (37.7%) are automatic backups from a later insurance verification screen. Interestingly, a further 3.9% of backups to this screen are from a question on whether the respondent is familiar with the family's health care coverage, which comes *after* the question on kinds of health insurance. Backups to this item average almost one every three interviews, which, given that this problem only occurred in the last two versions, likely underestimates the extent of this problem.

Another frequent target of backups is a question on sources of income. Following a series of four items in which the interviewer enters [1] for Yes or [2] for No, the fifth item (on Supplemental Security Income, or SSI), has the following response options: [1] Yes, the entire family, [2] Yes, some people but not everybody, and [3] No. Despite a warning to the interviewer in bold and upper case at the bottom of the screen ("PLEASE NOTE FIRST RESPONSE COVERS ENTIRE FAMILY, SECOND COVERS INDIVIDUAL FAMILY MEMBERS"), there are still 1,106 backups to this item, or an average of 1 in every 11 interviews. Of these, 87.4% are backups from a followup question asked of the [2]s only (asking which family members received such income); 97% of these backups returned to the SSI question to change a [2] to a [3].

These two examples are clear illustrations of inconsistencies in the instrument producing interviewer

difficulties (see Couper, 1994). Fortunately, in each case the problem was detected by the interviewer (because of the system not accepting a numeric response in the mark all that apply screen, and because of an inappropriate followup question in the SSI screen). We do not know about instances where the error was not detected by the interviewer.

A third item with a relatively large number of backups is a question on date of birth (about one backup per 17 interviews). Almost half of the backups to this question (45.3%) are from a subsequent age verification screen. Further inspection of the instrument reveals that if the interviewer enters that the age is incorrect on the verification screen, they are not taken back to the date of birth screen, but simply asked to record the correct age. We speculate that interviewers are realizing at this point that they have simply miskeyed the year of birth, and are using [F1] to backup and correct it on the previous screen.

The trace files also allow us to examine what the result of the backup was. For example, a frequent target of backups is an item in the household roster that adds persons. Every time this question is answered in the affirmative, an additional person loop is created. Of the 900 backups to this item, just over half (54.1%) are automatic backups from the household verification screen (described earlier in the context of [Sh-F6] use). In 61.4% of these automatic backups, an additional person was added. A further 33.8% of the 900 backups to this item are interviewer initiated, from a following question asking the name of the next person; of these., almost all (93.1%) result in changing a "yes" to a "no" suggesting the interviewer or respondent changed their mind about adding a person, or entered the wrong response. In sum, whether automatic or interviewer-initiated, 38.7% of all backups to the additional persons question result in an addition, 28.6% a change to '2' (no more persons), and the remaining 32.8% resulted in no change to the response on this item. This illustrates the importance of giving interviewers the flexibility to go back and change an answer, or simply to review the answer already there.

Finally, we also examined interviewer variation in use of various CAPI functions, using interviewer responses to background questionnaires administered during training. We find some evidence of an overall decline in use of various functions over time, suggesting that interviewers find these functions increasingly less useful, or increasingly forget (how) to use them. We also find that interviewers with less CAPI and survey experience appear more likely to enter notes (using [F7]). There is also less use of the jump functions (e.g., [F3] and [F4]) among those with more survey experience. We find few other interpretable trends, but are continuing the analysis

of interviewer correlates.

Summary

In this paper we have illustrated some of the uses of trace files in identifying problems with a survey instrument, and in evaluating interviewer knowledge and use of various CAPI functions. CASES trace files are somewhat limited in that they (1) don't reveal within-question editing and corrections, (2) only record actions executed by the system (i.e, miss function-key presses that produce no action), and (3) make it difficult to distinguish between interviewer and system actions. However, they are still a useful means of identifying what functions are being used by interviewers in the field, and where and how such functions are being used. Such knowledge could lead to changes in the design of the instrument, or improvements in interviewer training leading to more effective use of CAPI instruments. As noted earlier, when used in combination with other methods such as behavior coding or usability testing, trace file analysis is a powerful diagnostic tool in the instrument designer's toolbox.

References

- Baker, R.P. (1992), "New Technology in Survey Research: Computer-Assisted Personal Interviewing," *Social Science Computer Review*, 10, pp. 145-157.
- Couper, M.P., and Burt, G. (1994), "Interviewer Attitudes Toward Computer-Assisted Personal Interviewing (CAPI)," *Social Science Computer Review*, 12 (1): 38-54.
- Couper, M.P., Hansen, S.E., and Sadosky, S.A. (1997), "Evaluating Interviewer Use of CAPI Technology," in L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds), *Survey Measurement and Process Quality*, New York: Wiley.
- Couper, M.P., Sadosky, S., and Hansen, S.E. (1994), "Measuring Interviewer Behavior Using CAPI," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Alexandria: ASA, pp. 845-850.
- Hansen, S.E., Fuchs, M., and Couper, M.P. (1997), "CAI Instrument Usability Testing," Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Norfolk, VA, May.
- Sperry, S., Edwards, B., Dulaney, R., and Potter, D.E.B. (1996), "Human Computer Interaction and CAPI Navigation." paper presented at the International Conference on Computer Assisted Survey Information Collection, San Antonio, TX, December.

Table 1. Use of Function Keys

	Used by interviewer	Used in interview	Average per interview
[F1] Back up one item	99.7%	73.9%	13.56
[F2] Go forward one item	77.2%	9.6%	0.45
[F3] Jump forward	87.0%	17.5%	0.54
[F4] Jump table: this entity	51.8%	2.9%	0.08
[F5] Jump table: all entities	45.3%	2.6%	0.14
[F6] Change answer	46.6%	2.5%	0.08
[F7] Question-specific notes	61.8%	8.6%	0.18
[H] Access help screen	82.1%	9.4%	0.14
[Sh-F5] Language toggle	22.5%	0.75%	0.02
[Sh-F6] Window toggle	35.2%	1.7%	0.09
[Sh-F7] Show question specific notes	29.3%	1.1%	0.03
[Sh-F10] Show function key assignments	11.4%	0.3%	0.01

Figure 1. Household Roster Verification Screen

```

Caseid: 004
Item: MISPERS@MCHILD

-MISPERS- FR: READ FIRST TIME ONLY: I have listed as living here (READ NAMES).
PRESS "SHIFT-F6" TO SWITCH WINDOWS.

Have I missed -- (1) Yes (2) No (H)
- Any babies or small children? █
- Any lodgers, boarders or persons you employ who live here? █
- Anyone who USUALLY lives here but is now away from home traveling or in a hospital? █
- Anyone else staying here? █

HOUSEHOLD ROSTER
LINE HHSTAT NAME FX
-----
01 P John Smith 1
02 Jane Smith 1
03 Mary Smith 1
"PgDn = BOTTOM of screen " for next page
    
```