MEASURING SURVEY QUALITY IN A CASIC ENVIRONMENT

Mick P. Couper, Survey Research Center, P.O. Box 1248, University of Michigan, Ann Arbor MI 48106

Key Words: Computer assisted interviewing, data quality, evaluation

1. Introduction

Computer assisted survey information collection (CASIC) is now the default method of data collection for large government, academic and commercial survey organizations in the U.S. and Europe, and the use of computer assisted methods for a wide variety of applications is spreading rapidly. However, progress has been uneven in the transition from paper-and pencil data collection to computer assisted methods. While much progress has been made on technology issues (hardware and software), the field has been hard-pressed to keep up with the changes on the human side of the transition. This means changing the skills, tasks and tools of users to make optimal use of the new technologies. This also means changing the processes to make optimal use of the new technology.

A central belief of the CASIC perspective is acceptance of computer technology as an integral part of the entire survey process. This is not just data collection (where we have made much progress in recent years), but also processing, dissemination, management, etc. We are coming to realize that CASIC facilitates an integrated view of data collection and processing. Previous distinctions between different stages of the survey process are now fading. As a result of the adoption of new processes and procedures, the management and evaluation of the survey process also changes. The ability of CASIC methods to provide a wealth of real-time data on the progress of a study needs to be harnessed and managed to serve the goals of efficiency and quality. This will require organizational changes, retraining, new hardware and software systems, and so on (see Couper and Nicholls, 1998).

CASIC has thus brought changes in the work flow of survey activities. Traditional survey research was very much a batch-oriented activity. Questionnaires or interview guides were reproduced in batches, address labels or telephone numbers were assigned in batches, returned questionnaires were reviewed and keyed in batches, and the data were then computer edited and imputed in batches. With some exceptions (e.g., nonresponse reduction), this meant that quality indicators only became available after completion of a stage or batch, making informed intervention difficult. This *post hoc* approach to quality assessment limited the potential use of quality indicators in the ongoing survey management process, other than for repeated surveys. In an automated environment based on interactive computing, many activities are now done on a flow basis as sample cases are worked. This allows early detection of potential problems, and permits intervention when needed. However, the management of the survey process has not yet been sufficiently reengineered to take advantage of the new work flow, and to utilize the information that can be obtained to affect the process during the course of production.

This paper focuses on the auxiliary tools used both to measure survey quality in a production environment, and to manage the production of survey data to optimize quality and minimize costs. The particular focus is on the data collection process, and especially on computer assisted interviewing (CAI). However, much of what is written should apply to other parts of the process (e.g., data editing) or to other modes of CASIC data collection (e.g., CSAQ methods such as Web, TDE or VRE data collection). The central argument of this paper is that CASIC presents new challenges for the measurement and management of survey quality, but at the same time offers new tools and approaches for the effective evaluation of the process.

Surveys, and particularly large-scale ongoing government surveys, can best be described as production processes. In viewing them in this way, the importance of continuous quality improvement (CQI) or total quality management (TQM) can be stressed. Measurement is a key element in these perspectives. For example, Deming's 14 points on quality (cited in Dippo, 1998, p. 470) include the phrases, "...require statistical evidence that quality is built in..." and "...depend on meaningful measures of quality, along with price." Evaluation of quality is an ongoing part of the survey process that occurs (or should occur) before, during and after each round of data collection.

Quality measures can be viewed as both a management and evaluation tool. This is akin to Groves' (1989) notion of measurement and reduction. The quality and cost indicators collected on the survey process can and should serve both those wishing to measure the quality of a given data collection effort, and those attempting to minimize mean square error for a given cost. I take a broad view of quality in this paper, to include not only the traditional concepts of an absence of error or minimum mean square error, but also other aspects of process quality. The term "process quality" is used by Lyberg et al. (1997) in their volume to refer to this broader view of survey quality (see also Lyberg, Biemer, and Japec, 1998). Process quality includes such elements as cost efficiency, timeliness and relevance.

CASIC has not only changed the process being evaluated, but it has also changed the evaluation process itself. Some activities and quality indicators remain unchanged, regardless of the method of data collection. These include, for example, response rate measures (by interviewer, by region, etc.), and many elements of interviewer behavior (reading questions as worded, offering neutral feedback, probing, recording answers accurately, etc.). However, other evaluation criteria have become obsolete in an automated environment. For example, paper questionnaires were often inspected for completeness, legibility, skip pattern errors, and so on. This is no longer necessary in a CASIC environment. If correctly programmed, the CAI system will eliminate skip errors, not permit an item to be omitted without explanation or a specific missing data code (e.g., "don't know" or "refused"), and provide typed entries to open-ended questions for coders (although spelling errors are of course not eliminated). Furthermore, the introduction of CAI demands new skills on the part of interviewers (in CATI and CAPI) or respondents (in CASI or CSAQ), and evaluation methods need to be developed to see how well the new tools are being used.

At the same time as making some evaluation criteria obsolete, changing others, and adding new criteria, CASIC also changes the way we evaluate the process. It is thus important to review what we need to measure and how we need to measure it. In this paper I review the tools traditionally used to evaluate the quality of the data collection process and of the resultant data. I discuss the advantages and disadvantages of each of these methods. I examine how these are affected by the introduction of CASIC. I offer some thoughts on how technology can be harnessed to facilitate the quality measurement process.

This paper espouses a philosophy, rather than describing a situation already in existence. While many survey organizations already do (or can quite easily) collect the kind of data I describe here, we as a field are far from having all the tools to manipulate, analyze, and otherwise turn these data into useful information. I believe the development of such tools is imminent, and look forward to the day when these are a routine part of the data collection process itself. Until then, we have much to learn about what data are most useful for evaluating the process, and what kinds of information are needed by which people to manage and assess various parts of the process.

Others have eloquently articulated the importance of quality measurement as an overall strategy in surveys (e.g., Biemer and Caspar, 1994; Lyberg, Biemer, and Japec, 1998; Morganstein and Marker, 1997). My goal is to offer some observations on the methods that are currently available or soon will be that will facilitate the evaluation of the survey process in a CASIC environment. This paper focuses primarily on the evaluation of the data collection process, but CASIC also affects the design and development process and back-end processing. These two areas are also covered briefly.

2. Instrument Design and Development

In recent years a large number of tools have been developed and implemented to evaluate survey instruments during the design stage. Many of these involve some or other form of cognitive laboratory testing (see Forsyth and Lessler, 1991) and a conventional pretest with monitoring (in telephone surveys) or behavior coding of tape recorded interviews (in personal visit surveys). These methods provide a wealth of qualitative and quantitative data on respondents' interpretation and understanding of survey questions, their reactions to the questions, and the responses they provide. These methods also reveal difficulties interviewers may have reading the questions, selecting appropriate response alternatives, choosing appropriate probes, and so on. These all serve as potential indicators of problems with question wording.

Computer assisted interviewing (CAI) has introduced additional complexities to the task. Bethlehem (1997, p. 377) asserts that "Computer assisted interviewing makes the work of the interviewers easier. Since the computer is in charge of determining the proper route through the questionnaire, the interviewers can concentrate on asking the questions." This is an unsubstantiated claim, and tends to receive little support when observing CAI interviews in a usability laboratory. There are several reasons for this, I believe. First, reading questions and following skips is only one part of an interviewer's task — other aspects of interacting with a computer to change answers, review previous items, etc., are required. Second, with the advent of CAI, the complexity of survey instruments has increased dramatically. Third, while the design of paper instruments has developed over several decades and their form and function are familiar to interviewers, relatively little attention has been paid to the design of CAI instruments (see Couper, 1999). Many CAI designs I have seen make it hard for the interviewer to find the question to be read and to determine what to do. This suggests it is important to understand how interviewers interact with the survey instrument, and to examine the kinds of difficulties they experience when doing so. In our work evaluating the usability of survey instruments (Hansen, Couper, and Fuchs, 1998), we have found that interviewer difficulties with the CAI instrument interfere at times with the interviewer-respondent interaction.

In recent work (Couper and Schlegel, 1998; Hansen, Couper, and Fuchs, 1998; Lepkowski et al., 1998) we have sought to combine three methods to evaluate the quality of survey instruments in CAI. These are behavior coding, usability testing and trace file analysis. We have found that behavior coding is a useful tool for revealing problems of comprehension and communication, particularly in the interviewer-respondent interaction. Trace files (or keystroke files) provide information on the interviewer-computer interaction, focusing on the interviewer input. Usability testing provides the most comprehensive information on the entire process, both the interviewer-respondent and interviewer-computer interactions, as well as their effect on each other. While they yield rich qualitative data, usability evaluations may be costly and time-consuming to conduct. Given the relative costs and effort involved in these three approaches, we recommend a combination of methods be used during pretesting to evaluate all aspects of the survey instrument (see Lepkowski et al., 1998).

3. Data Collection Process

A variety of tools have been used to evaluate the quality and costs of an ongoing survey operation. These include tape recording and behavior coding (for field and telephone interviews), monitoring (in centralized telephone facilities), reinterviews, review of completed questionnaires, and performance and production measures. These are all methods that can be used throughout the field period for continuous quality improvement. In addition, there are other periodic evaluation methods such as mock interviews, tests of interviewer knowledge or practice, and observation by supervisors, that can be used to certify interviewers following training, or to evaluate interviewer performance at set points during the survey. We discuss each of the methods briefly in turn, noting the effect that CASIC has had on their use.

3.1 Tape Recording and Behavior Coding

A traditional approach to evaluating the quality of several aspects of the survey process is to have interviewers tape a random subset of their interviews in the field, and have the interview behavior coded (see Cannell, Fowler and Marquis, 1968; Fowler and Cannell, 1997; Lepkowski et al., 1998; Mathiowetz and Cannell, 1980). This approach is both expensive and time consuming, and subject to many errors (interviewers may forget to record, the tape may be inaudible, etc.). Furthermore, the method is intrusive — interviewers and respondents know they are being recorded and could change their behavior as a result. The coding process is also lengthy and inefficient, and the data are usually only available several months after collecting the tapes.

An alternative approach under CASIC would be to have the computer record a subset of items for later coding. This approach has several advantages. First, it is less intrusive than setting up a tape recorder prior to the interview. All respondents could be informed of the possibility of taping, but interviewers would not know which items were being recorded. Second, one can target certain questions (e.g., all open-ends), or record a random subset of items (e.g., every 10th item). This approach is reminiscent of the random probe technique proposed many years ago by Schuman (1966), but never implemented. We now have the wherewithal to automate this process.

Because these data are in digitized form, they can be much more easily sorted, rearranged, subsampled, etc., than was possible using the physical and sequential medium of cassette tapes. Digitized sound files are easily transmitted and stored, further speeding up the coding process. We suspect that coding efficiency and reliability would be improved if coders were able to code all instances of a single exchange (question-level interaction) before proceeding to the next question. Software already exists for the analysis of such digitized sound files (e.g., Dijkstra's (1995) Sequence Viewer). Selected segments could easily be identified for more detailed analysis, for example using conversational analytic methods (e.g., Schaefer, 1990; Schaefer, Maynard and Cradock, 1993). The files can also be subject to automatic transcription (using intelligent voice recognition software) to facilitate the coding process.

The technology is available to do this now. We suspect this approach will be implemented in CASIC surveys within the next several years. Concerns about degradation of system performance will be overcome as more powerful processors become available. Probably the biggest challenge is the issue of informed consent and confidentiality. One could imagine overcoming this by having a standard statement read at the beginning of each interview to obtain permission for taping. Another challenge is that such a system has the capacity to generate large amounts of data that may never be coded. Thus, this approach should be thoroughly evaluated and used with due caution.

3.2 Monitoring

Monitoring of telephone interviewing is a technique that was developed largely in the era of CATI, so already makes use of many of the automated features of a centralized CATI system. In the early days, the selection of interviewers to monitor was left to the judgement of the monitor or supervisor. Since then, developments have led to sampling schemes for selecting interviews or portions of interviews to listen to (see Couper, Holland, and Groves, 1992). Early inefficiencies in searching for appropriate interactions can now be eliminated, by using the CATI software to detect where each interviewer is in the interview, and directing those cases to the monitor. Similarly, early monitoring reports were based on subjective evaluations of various interviewer behaviors; these have been replaced with more objective evaluation criteria based on behavior coding schemes. Still more recently, the recording of the monitoring evaluations is done directly on computer, obviating the need for an additional keying step. Recent developments have also led to the development of statistical process control charts based on monitoring data, with systematic procedures for feedback and evaluation (see Mudryk, Burgess and Xiao, 1996). Thus, CASIC permits several enhancements and extensions to existing evaluation systems based on monitoring.

3.3 Reinterviews

Reinterviews are a costly evaluation tool, but very useful for estimating response variance and response bias, and detecting interviewer falsification (see Forsman and Schreiner, 1991). CAI has not eliminated the need for reinterviewing, but has made the process easier. Both survey and production data are available from the field much sooner, permitting adjustments to the sampling scheme to maximize the efficiency of the reinterview operation. For example, sample cases with unusual distributions on key variables, with extreme values of interview length, with very high or very low proportions of don't knows or refused item — in short, anything from the production and performance statistics that indicates a potential problem — can be sampled at a higher rate for reinterview. This also means that the reinterview can be completed closer to the time of the original interview, minimizing differences due to recall problems or to true changes in respondents' answers.

Another issue in reinterviews has been whether to blind the second interviewer to the responses obtained by the original interviewer. This avoids contamination, but also makes clarification of any resulting discrepancies with the respondent difficult. CAI systems can be programmed to compare automatically the two responses, and reveal them to the second interviewer only if a discrepancy occurs (see Brown, Hale, and Michaud's (1998) discussion on reactive feedback). Thus, the reconciliation process can be facilitated by CAI, providing the ability to conduct independent interviewing with immediate verification. Depending on the purposes of the reinterview program, these features of CASIC can be used to make the process more efficient.

3.4 Performance and Production Measures

Let's take the example of a personal visit survey. In the "old days" of paper-and-pencil, interviewers mailed completed questionnaires to a supervisor, along with a completed coversheet for that case. These data would be reviewed by the supervisor, and may have been keyed for later analysis. For more up-to-date information on the progress of a study, interviewers would file regular reports by mail or telephone with their supervisors. These reports would then be combined and sent to head office where an overall summary of the process to date was obtained. Given the labor-intensive nature of this task, usually only a small number of items were reported, and the summary reports were several days or weeks old by the time they were given to the project managers. These reports may contain a summary of the status of each sample case (interviewed, refusal, noncontact, etc.). Often separate time and expense reports were submitted by interviewers on a regular basis, permitting some estimation of the costs expended to date on the project. Supervisors would often review completed paper questionnaires, and use indicators such as the legibility, length or clarity of the open-ended responses, the completeness of the questionnaire (number of missing items), skip errors, etc., to evaluate the interviewers' work.

What do things look like under CASIC? First, completed instruments are transmitted directly to head office on a regular basis (often nightly). The data are already in electronic form and can thus be tabulated instantly to provide project managers with a daily update of the progress of the study. A variety of measures are available directly from the automated instrument. These include the substantive data themselves (permitting detection of outliers and strange patterns), counts of "don't knows" and refusals, changed answers, edit failures, and a variety of timer information.

Second, the case management system is also automated under CASIC, replacing paper coversheets. This means that in addition to substantive data being instantly available for evaluation and review, a variety of process data are also available. Supervisors can review the status of any case at any time. The number of calls or visits — and the timing of those calls — to a particular sample unit or cluster can be ascertained in sufficient time to give direction to field staff about follow up efforts, to permit projection of likely response rate goals, and to allocate appropriate resources to meet those goals.

Third, in many organizations, time and expense reports are now also automated, with interviewers submitting this information on a flow basis in electronic form. This again means that cost data can be made available almost instantly to staff to assess progress of the study. Thus, we may know not only how many cases a particular interviewer still has outstanding, but we can also estimate the cost of obtaining an interview from each of those remaining cases.

These data are all available virtually in real-time. Project managers can view daily progress reports on the entire process, or some subset of it, at any time. Data are available electronically, allowing managers to manipulate the information as they desire (if they have the right skills and tools to make use of the information available). Clients and remote personnel (e.g., regional office staff) can have access to the production data through wide area networks or Internet links (see Smith, Rhoads, and Shepherd, 1998).

While the performance and production measures are primarily used to support the management of the data collection process, they also provide a wealth of information to facilitate the quality assessment process. However, there are a number of drawbacks to the use of these various data sources. First, these tools can provide too much data, overwhelming the capacity of managers and evaluators to handle the information. Many of these reports are still delivered to managers in the form of printed output, containing large amounts of data that are difficult to digest and use for statistical analysis. We need tools that permit managers and evaluators to query these databases to obtain a variety of generic and customized reports, and obtain charts and other graphics that will allow them to identify processes out of control, or areas that require intervention.

Second, the level of detail that is available is often dictated by the quality of the software used. For example, some case management systems operate on a case level. Each sample case is a record, and only aggregate information such as the total number of calls or the time and outcome of the final call are available. Other systems are designed to track individual calls, providing much more detail on every attempt made to contact each sample household. Possible measurement problems may arise to affect the quality of these data; for example, interviewers may not switch the computer on for every visit, or not until they enter the house, or they may not close out the interview until leaving the household. Often, the data that are available from the case management system reflect the particular design and organizational philosophy behind the development of the system. The production and performance data may be of poor quality because they are designed for other purposes and because nobody has taken the time to evaluate their utility for quality measurement. Often simple programming changes will produce data more usable for both purposes.

3.5 Keystroke Files

Keystroke files (also called trace files or audit trails) are automatic byproducts of many CAI software systems. These are essentially a record of every key pressed by an interviewer (or respondent) as he/she moves through the instrument. These files vary in the amount of information captured, and hence in their utility for quality measurement. For example, the trace files produced by the CASES system record only those interviewer actions recognized and acted on by the software. Surveycraft's keystroke files, on the other hand, capture every single key pressed by an interviewer, whether or not that key was valid for the software application. Using these latter files we have examined the use of erroneous or invalid keys and explored the inefficiencies of interviewers in correcting open-ended text entries (see Couper, Hansen and Sadosky, 1997). We have also used keystroke files to examine difficulties interviewers may have with certain items (see Couper and Schlegel, 1998) and to evaluate respondent performance in a CASI interview (see Caspar and Couper, 1997).

Keystroke files were originally designed for debugging purposes, to permit replay of the instrument. As such, they were relatively unsophisticated, and difficult to use for evaluation purposes. These have been enhanced over the last few years to make them much more amenable to inspection and analysis. For example, Surveycraft's keystroke files now also capture the question or item number, permitting item-level analysis. Furthermore, both the CASES and Surveycraft systems can be extended to include item-level timers.

Despite their easy availability, keystroke files have a number of disadvantages. First, their sheer number and unstructured format may quickly overwhelm the analyst. This has generally precluded their use for purposes other than testing or debugging. Much of our early work has focused on finding ways to summarize the files and extract the pertinent information to facilitate analysis (see Couper, Hansen, and Sadosky, 1997). Second, keystroke files capture only one part of the interaction in a CAI interview — that between the interviewer and computer; they provide little or no information on the interviewer-respondent interaction. They also do not reveal what the interviewer intended to do on a particular item; this must be gleaned from an examination of the individual keystroke files. Nonetheless, we have found keystroke files very useful as a supplement to other methods of interviewer or instrument evaluation.

Another potential problem with keystroke files is that as CAI systems move to Windows, finding graphical user interface (GUI) equivalents of keystroke files may be hard. There are no generic programs available to do this, so these functions will need to be built into the CAI applications directly.

For Web surveys, a number of "user metrics" can be built into the instrument to evaluate the respondent's interaction with the Web instrument. These may include examination of points at which respondents abandon the survey, counts of missing items, backups to change answers, time per screen, etc. (Jeavons, 1996; Nichols and Sedivi, 1998).

3.6 Time Measures

In the traditional approach the interviewer records the start and end time of the interview on the paper questionnaire. Sometimes start and/or end times of sections of the instrument are also recorded. These data are then keyed along with the substantive data for the survey, and may be made available to analysts and managers with the full data set. These time measures are often inaccurate, prone to missing data, and available too late to affect the current production cycle.

CAI permits automated timers to be embedded in the instrument. These can be as detailed as are needed, from overall instrument times in minutes to numerous section-level times in hundredths of seconds.

Some CAI systems now permit the automatic recording of item-level timers along with keystroke files. Such measures are especially useful during the pretest phase of a survey, in order to identify particularly lengthy or burdensome items that could be considered candidates for exclusion. But item-level timers could also be used as an automated substitute for response latency measures (e.g., Bassili, 1993). We are currently engaged in a project to evaluate whether automatic timers embedded in the instrument can replace interviewer-activated response latency measures.

Item-level timers could also be used to identify interviewers who may be going at too fast a pace or not reading the full text of a question. This could be done by establishing benchmark (or "gold standard") times for particular items, and comparing average times to the benchmark. Caspar and Couper (1997) used this approach in the evaluation of an audio-CASI instrument. We found that, on average, ACASI respondents tend not to listen to the full question over the headphones before they provide an answer. Furthermore, the proportion of the question listened to declines over the course of an instrument.

This suggests that respondents increasingly speed up the answering process over successive sections of the interview. This finding has possible implications for the design of survey questions; for example, if qualifying information is placed at the end of the questions, respondents may be more likely to ignore the information, particularly later in the interview.

Examining trace files from the same study, Caspar and Couper (1997) also found that only a small percentage (4.7%) of audio-CASI respondents ever turned the screen off during the self-administered interview, and none turned it off for the entire time. Thus it appears that the audio mode is largely used as a supplement to the text mode, and is needed only when greater privacy is required. Thus, these measures appear to be equally useful in exploring respondent interaction with a CASI or CSAQ instrument as they are for studying interviewer behavior.

3.7 Tracking Data and Problem Reports

Many survey organizations maintain a help desk for CASIC users, or collect problem reports from the field. These include reports of hardware and software failures. These data are rarely (if ever) collated and used to evaluate the process. They could be useful in identifying interviewers with unusually high rates of problems, or to identify system components with high rates of failure. For example, many organizations assume a three-year replacement cycle for laptop computers in the field. This could be verified empirically using these data. This could inform decisions about replacement machines or the acquisition of backup components (e.g., batteries, hard drives). The frequency and severity of reported problems could be used to inform priorities for action. The information could also be used to prepare process control charts focusing on hardware and software failures in the field.

3.8 Interviewer Notes

Almost every CAI software system has the capacity for interviewers to record notes, either at the item level or at the overall instrument level. These take time and effort for interviewers to complete (often during the interaction with respondents), yet it seems that these notes are rarely reviewed or analyzed, probably because of the sheer number of notes generated from a large study. We are exploring the use of content analysis software (Nu*Dist) to classify these notes into meaningful groups (e.g., hardware, battery, screen, software, etc.). This will reduce the amount of information managers will need to peruse the assess the extent and severity of a particular type of problem. This sorting of the notes could then permit a manager to more easily review a set of notes about a particular issue. In addition, quantitative information (counts of problems) can be obtained to ascertain the relative frequency of various difficulties. Interviewer notes are thus another potentially useful source of data on the process, if we can find ways to reduce the amount of information produced.

4. Data Editing and Post-Interview Processing

This last area I cover only briefly, but it is no less important than the other areas of evaluation. Under the paper and pencil model, supervisors, editors, coders and other staff often used different colored pencils to make changes to the instrument. This way one could track who made what changes. However, usually only the final edited questionnaires were keyed, leaving the earlier versions in storage for review if needed. Thus, little quantitative information was available to evaluate the post-interview processing step, probably one reason why this part of the process has received less research attention than others.

In the CASIC model, many of these processes have now been automated. Much of the editing has been built into the survey instrument itself, and inconsistencies are resolved with the respondent during the interview. Keystroke analysis can reveal how much editing is being done in the interview and where it is occurring. In computer assisted editing and coding, the computer automatically keeps track of which changes were made and by whom, and usually also keeps the old values for reference. Coders can also be asked to provide reasons for a particular change, and this could also be part of the data record. Similarly, time measures can be automatically captured for the coding process.

Biemer and Caspar (1994) present data from an evaluation of a coding process. They provide documented evidence that continuous quality improvement does work in survey applications. Fred Conrad and I are conducting similar analyses of the discrepancy reports (from double-coding of a systematic sample of cases) from the industry and occupation process for the Current Population Survey to understand (and hopefully improve) the coding process.

5. Summary and Conclusions

In this paper I have reviewed some of the tools and techniques available for evaluation of the process of survey data collection. Many of the CASIC tools represent enhancements over similar tools used in paper and pencil surveys (e.g., reinterview, monitoring). Others are new tools unique to an automated data collection environment (e.g., keystroke files, item-level timers).

Regardless of whether new or old methods are used, the automation of the process makes it much easier to obtain data during the process itself, permitting intervention where needed. What CASIC has brought is the ability to measure many more aspects of the survey process than was previously feasible. Many of these measures can be passively collected, requiring no extra effort on the part of interviewers. The remaining challenges are to evaluate the utility of these measures as indicators of quality, and to find ways to make this information routinely available, both for quality assessment and for management of ongoing surveys.

CASIC has the capacity to overwhelm us with data on the survey process. We need automated survey management systems that allow us to track, measure and utilize the information. This would promote effective management of the process and well as provide timely data for evaluation of the quality of the process. Case management systems will likely form the foundations for such survey management software tools.

In the paper days often several weeks or months may pass after the completion of data collection before the data were available for inspection. Problems could only be corrected on the next wave of data collection. In a CASIC environment, we can track the progress of production virtually in real time, and have the ability to make corrections on the fly. We wouldn't want to do this often, but having access to timely information can help prevent a disaster, and provide the information necessary to make optimal decisions.

Despite the wealth of auxiliary data that are potentially available from CASIC surveys, we have yet to fully harness this information. There are several challenges that remain if we are to fully utilize the data that CASIC provides for ongoing quality assessment.

It is important that we use these data responsibly. First, if we collect the data, we should use it. Second, the information we obtain must be used to provide feedback to the production process. Third, as already noted, we must develop efficient ways to access, summarize, analyze and store the data. Finally, we must behave responsibly towards those who provide the data (whether they do so knowingly or not). Using the process data without informing interviewers and/or respondents may lead to suspicions and a sense of being "watched." In addition, the feedback we provide communicates to interviewers (and other staff) the relative importance of various components of the process. For example, focusing primarily on response rates may suggest that data quality is less important. We need to be careful about the message that the evaluation process conveys to those being evaluated.

In sum, CASIC offers us much potential to make progress on continuous quality improvement measures and practices. It remains to explore the optimal combination of methods and identify the most applicable information, and to build the collection and analysis of these data into the ongoing survey process.

References

Bassili, J.N. (1993), "Response Latency Versus Certainty as Indexes of the Strength of Voting Intentions in a CATI Survey," *Public Opinion Quarterly*, 57 (1): 54-61.

Bethlehem, J. (1997), "Automated Control Systems for Survey Processing," In Lyberg, L.E., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (1997), *Survey Measurement and Process Quality*. New York; Wiley.

Biemer, P. and Caspar, R. (1994), "Continuous Quality Improvement for Survey Operations: Some General Principles and Applications," *Journal of Official Statistics*, 10 (3): 307-326.

Brown, A., Hale, A., and Michaud, S. (1998), "Use of Computer Assisted Interviewing in Longitudinal Surveys," In Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls, W.L. II, and O'Reilly, J. (eds.) (1998), *Computer Assisted Survey Information Collection*. New York: Wiley.

Cannell, C.F., Fowler, F.J. and Marquis, K.H. (1968), "The Influence of Interviewer and Respondent Psychological and Behavioral Variables in the Reporting in Household Interviews," *Vital and Health Statistics*, Series 2, No. 26, Washington, DC: U.S. Government Printing Office.

Caspar, R.A. and Couper, M.P. (1997), "Using Keystroke Files to Assess Respondent Difficulties with an ACASI Instrument." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Anaheim, CA, August.

Couper, M.P. (1999), "The Application of Cognitive Science to Computer Assisted Interviewing." In M.G. Sirken, D.J. Hermann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (eds.). *Cognitive and Survey Methods Research*. New York: Wiley, forthcoming.

Couper, M.P., Holland, L., and Groves, R.M. (1992), "Developing Systematic Procedures for Monitoring in a Centralized Telephone Facility," *Journal of Official Statistics*, 8 (1): 63-76.

Couper, M.P. and Nicholls, W.L. II (1998), "The History and Development of Computer Assisted Survey Information Collection." In Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls, W.L. II, and O'Reilly, J. (eds.) (1998), *Computer Assisted Survey Information Collection*. New York: Wiley.

Couper, M.P. and Schlegel, J. (1998), "Evaluating the NHIS CAPI Instrument Using Trace Files." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Louis, May.

Dijkstra, W. (1995), "SEQUENCE: Computer-Assisted Coding and Analyzing Questions and Answers in Survey Interviews." Paper presented at the Workshop on Interviewer-Respondent Interaction in the Standardized Survey Interview, Amsterdam, November.

Dippo, C. (1998), "Survey Measurement And Process Improvement: Concepts and Integration," In Lyberg, L.E., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (1997), *Survey Measurement and Process Quality*. New York; Wiley.

Forsman, G. and Schreiner, I. (1991), "The Design and Analysis of Reinterview: An Overview," In Biemer, P.P., Groves, R.M., Lyberg. L.E., Mathiowetz, N.A. and Sudman, S. (eds.), *Measurement Errors in Surveys*. New York: Wiley.

Forsyth, B.H. and Lessler, J.T. (1991), "Cognitive Laboratory Methods: A Taxonomy," In Biemer, P.P., Groves, R.M., Lyberg. L.E., Mathiowetz, N.A. and Sudman, S. (eds.), *Measurement Errors in Surveys*. New York: Wiley.

Fowler, F.J. and Cannell, C.F. (1997), "Using Behavior Coding to Identify Cognitive Problems with Survey Questions," In Schwarz, N., and Sudman, S. (eds.), *Answering Questions*. San Francisco: Jossey-Bass.

Groves, R.M., (1989), *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R.M. and Tortora, R.D. (1991), "Developing a System of Indicators for Unmeasured Survey Quality Components," *Bulletin of the International Statistical Institute*, 48th Session, Volume 2, pp. 469-486.

Hansen, S.E., Couper, M.P., and Fuchs, M. (1998), "Usability Eevaluation of the NHIS instrument." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Louis, May.

Jeavons, A. (1996), "Trends in Interviewing." Paper presented at the MRSA Conference, Gold Coast, Australia.

Lepkowski, J.M., Couper, M.P., Hansen, S.E., Landers, W., McGonagle, K.A., Schlegel, J., Wright, T., and Chevarley, F. (1998), "CAPI Instrument Evaluation: Behavior Coding, Trace Files, and Usability Methods." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Louis, May.

Lyberg, L.E., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (1997), *Survey Measurement and Process Quality*. New York; Wiley.

Lyberg, L.E., Biemer, P., and Japec, L. (1998), "Quality Improvement in Surveys — A Process Perspective," paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, August.

Mathiowetz, N.A. and Cannell, C.F. (1980), "Coding Interviewer Behavior as a Method of Evaluating Performance," *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 525-528.

Morganstein, D. and Marker, D.A, (1997), "Continuous Quality Improvements in Statistical Agencies," In Lyberg, L.E., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (1997), *Survey Measurement and Process Quality*. New York: Wiley.

Mudryk, W., Burgess, M.J., and Xiao, P. (1996), "Quality Control of CATI Operations in Statistics Canada," *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 150-159.

Nichols, E. and Sedivi, B. (1998), "Economic Data Collection Via the Web: A Census Bureau Case Study." Paper presented at the Joint Statistical Meetings of the american Statistical Association, Dallas, August.

Schaeffer, N.C. (1990), "Conversation with a Purpose — Or Conversation? Interaction in the Standardized Interview," In Biemer, P.P., Groves, R.M., Lyberg. L.E., Mathiowetz, N.A. and Sudman, S. (eds.), *Measurement Errors in Surveys*. New York: Wiley.

Schaeffer, N.C., Maynard, D.W. and Cradock, R.M. (1993), "Negotiating Certainty: Uncertainty Proposals and Their Disposal in Standardized Interviewing," Madison: University of Wisconsin, Center for Demography and Ecology (Working paper no. 93-25).

Schuman, H. (1966), "The Random Probe: A Technique for Evaluating the Validity of Closed Questions," *American Sociological Review*, 31 (2): 218-222.

Smith, J., Rhoads, M., and Shepherd, J. (1998), "Getting from There to Here: Electronic Data Communications in Field Surveys," In Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls, W.L. II, and O'Reilly, J. (eds.) (1998), *Computer Assisted Survey Information Collection*. New York: Wiley.