THE FEASIBILITY OF MONITORING FIELD INTERVIEW PERFORMANCE USING COMPUTER AUDIO RECORDED INTERVIEWING (CARI)

Paul Biemer, Deborah Herget, Jeremy Morton, and Gordon Willis (Research Triangle Institute)
Paul Biemer, Research Triangle Institute, P.O. Box 12194, RTP, NC 27709

Key Words: nonsampling error; quality control; CAPI; reinterview; survey costs

1. Introduction

Computer Audio Recorded Interviewing (CARI) is a computer software application, developed by RTI computer scientists, which allows the computer to act as a sophisticated tape recorder as the interviewer administers a CAPI (Computer Assisted Personal Interview) questionnaire. CARI unobtrusively digitally records the audio verbal exchange between the interviewer and respondent during production interviews. The system is completely under software control such that at any predetermined or randomly selected point during the interview, the recording can begin or terminate. This function provides a capability which has previously been unavailable in a field interviewing environment: the capacity to provide an audio record of the exchange for purposes of interview monitoring, similar to that routinely used in centralized CATI telephone facilities (see Couper, Holland, & Groves, 1992; Massagli & Clarridge, 1995; Stanley & Erth, 1995.

CARI potentially meets several critical needs specific to field interviewing. In general, it provides a means for monitoring the quality of the field interview, including the behavior of the interviewer during the interview and the reactions of the respondent to survey questions (see Mudryk, Burgess, & Xiao, 1996, for a review of issues related to survey quality control).

CARI can potentially be used for a range of applications, including:

- Detecting gross departures from appropriate procedures, such as interview fabrication;
- Evaluating interviewer execution of interviewing guidelines with the intent of providing both positive and corrective feedback,
- Identifying questionnaire problems and data collection difficulties using interviewer-respondent interaction coding.
- Collecting verbatim responses to open ended questions in an interview.

RTI and the Census Bureau are collaborating on a study to determine the basic feasibility of CARI for production CAPI surveys. Our investigations are designed to address the following research questions:

1. What is the audio quality of CARI recordings? Can both the interviewer and respondent be heard clearly and distinctly?
2. Are respondents cooperative with the request to allow CARI to record during interviews? What proportion of survey respondents refuse the request?
3. How do field interviewers react to CARI? Do they recognize and appreciate the potential of CARI to enhance data quality?
4. How does CARI affect CAPI system performance? Is there any noticeable degradation of system response time or other tell-tale signs that might alert the interviewer as to when CARI is recording?
5. How does the cost of a CARI-based interviewer verification system compare to the costs of traditional reinterview/recontact approaches for verification?
6. In general, is CARI operationally feasible for production surveys? What issues remain before CARI can be recommended for implementation in Census Bureau surveys?

To answer these questions, RTI implemented a CARI-based interviewer verification system in the National Survey of Child and Adolescent Well-being (NSCAW) and conducted a series of evaluations to assess system performance. The NSCAW is a panel survey conducted by RTI of 6,700 children who are subjects of abuse and neglect reports. The survey was fielded in September 1999 and will continue until March 2003. An overview of the CARI-based procedures that have been implemented in the NSCAW is provided in the next section.

The purpose of this paper is report on the results we have accumulated to date on the feasibility questions posed above. After describing the NSCAW implementation of CARI, we provide a summary of the feasibility results in Section 3 and conclude with a description of future directions of the research.

2. Implementing CARI for the NSCAW

Description of the NSCAW. In 1996, Congress directed the Secretary of the U.S. Department of Health and Human Services (DHHS) to conduct a national study of children in the child welfare system as well as other children in the U.S. who are at risk of abuse or neglect. The Congress directed that the study follow children over time and collect data on the types of abuse or neglect involved, their agency contacts and services, any out-of-home placements (such as foster care). The directive

mandated that the survey yield reliable estimates at the state-level for as many states as possible. The Administration on Children, Youth and Families (ACYF) has undertaken the National Survey of Child and Adolescent Well-Being (NSCAW).

During the first 12 months of data collection, face-to-face, CAPI baseline interviews will be conducted with 6,700 children who have been investigated for child abuse and neglect by the local child welfare agency. Follow-up interviews will alternate by telephone and face to face at six month interviews. Survey respondents include children, their caregivers, the child welfare agency caseworker, and teachers. CARI has been implemented for the child, caregiver, and careworker CAPI interviews; however, our feasibility study will focus on the caseworker and caregiver interviews. The former is conducted in an office environment similar to many establishment and institutional surveys while the later is conducted in a home environment typical of many household surveys of low income populations.

Summary of CARI Procedures. For any large-scale field survey, the risk of interviewer fabrication is considerable and field verification procedures are needed (Biemer and Stokes, 1989). Because of the high percentage of nontelephone households in the NSCAW sample, traditional verification methods which rely on telephone contacts were not feasible in this study. Further, face-to-face verifications were deemed prohibitively expensive due to high travel costs. Since CARI provides a cost-effective means for verifying that an interview was conducted in both telephone and nontelephone households, detection and deterrence of fabricated interviews were the primary motivation for implementing CARI for the NSCAW.

However, CARI also provides the means for a continuous monitoring of interview quality, interviewer performance, questionnaire performance, the reactions of the respondent to survey questions. Thus, a secondary objective of the CARI implementation was to extend its use to incorporate these other interview quality improvement features.

CARI was introduced to all NSCAW interviewers during training, at which time the interviewers were briefed on the purposes and uses of CARI and trained in CARI system procedures. CARI training includes informed consent procedures, description of system operations, and procedures for downloading CARI digital audio (.wav) files. Due to the size of the CARI audio files, uploading the files to the RTI central computer via a phone modem was infeasible. Instead, a Zip disk shuttling system was devised for transferring the files. Each interviewer was equipped with a Zip drive and was instructed on how to connect the Zip drive to the laptop computer, copy the CARI files to a Zip disk, and transmit the CARI files to RTI by regular mail.

Interviewers were also trained thoroughly on the use of CARI in the field. Briefly, the procedures involve reading a short informed consent statement to the respondent prior to beginning the CAPI interview to obtain consent for the recording of the interview. Interviewers are instructed to emphasize that these recordings are used strictly for quality control purposes. The respondent is asked to sign the consent form as evidence that consent was either granted or denied. Then, at the start of the CAPI interview, the respondent's decision to either allow or refuse CARI recording is repeated by the interviewer followed by the question "Is that correct?" This interchange is the only one in the interview that is recorded by CARI if the respondent refuses to allow CARI recording. Otherwise, by the interviewer entering a code indicating consent for CARI recording, CARI is activated for the remainder of the interview. In this way, both written and oral consent of the respondent is obtained for the use of CARI during the interview.

Following the interview, the interviewers copy the CARI audio files onto Zip disks and mail them to the RTI data quality monitors in Research Triangle Park, NC. Each disk can hold all the audio files from approximately four interviews. Thus, the interviewers are instructed to copy their CARI files onto Zip disks once per week or after four interviews had been completed, whichever comes first. A menu item in the case management system is used for copying the audio files to the Zip disks.

When the CARI audio files are received in RTP, they are transferred to the CARI server and are then available to the q.c. (quality control) monitors. Although portions of all interviews are recorded, CARI monitors only listen to approximately 10% of each interviewer's recordings to authenticate the interview. Cases that raise the suspicions of the q.c. monitors are transferred to the telephone center for further follow-up via telephone contact of the respondents. In addition, approximately 10 percent of the respondents who refuse to allow CARI to record their interviews are also followed up in the telephone center.

Though the primary use of CARI is for interview verification, it also provides the unique opportunity to provide performance level to field interviewers. As needed, data quality or performance problems are forwarded to the interviewer's field supervisor so that the supervisor can address these performance issues. These include problems with question administration, general interviewing techniques, payment of incentives, and adherence to study procedures. Positive feedback is also provided to the field supervisors to be shared with the interviewers either individually or in group sessions.

This feedback may regard interviewing techniques (question delivery, probing, feedback, etc.), handling difficult situations, and answering respondent questions about the survey.

3. Results From the Feasibility Evaluation

The NSCAW is the first national study using CARI for verification and interview improvement purposes. Prior to implementing CARI for this survey, RTI had only conducted small field tests to ensure the general feasibility of CARI for large scale implementation. Therefore, a number of questions regarding CARI feasibility needed to be addressed in this implementation. Some of the areas we have evaluated as the system is being used in the NSCAW include: 1) potential problems with system performance; 2) audio quality of recorded interviews;
3) respondent acceptance of CARI; 4) interviewer and other project staff reactions to CARI, and 5) implementation costs. Our findings in each of these areas is discussed in the following.

Effect of CARI on Computer System Performance. To determine whether the CARI system interfered in any way with the CAPI interview, we tested the system extensively in the NSCAW pretest interviews. Then, in the field, we used the NSCAW problem identification system to document any problems with either the laptop computers or the CAPI software that could be attributed to the CARI system. We found no degradation of CAPI system response time or CAPI questionnaire performance. Further, no software or hardware problem that could be attributed to CARI has been reported so far in nine months of fieldwork. Indeed, the CARI system appears to be operate completely unobtrusively on the laptop computers.

CARI Audio Quality. A critical criterion for the feasibility of CARI for production use is audio quality. If the CARI monitors were unable to hear the respondent or the interviewer in a large proportion of interviews, use of CARI as a quality-control tool would be quite limited. Due to the position of the built-in microphone on most laptop computers, we expected that the interviewers would be quite audible in all but a few cases. However, there might be many situations where the respondent is inaudible due to background noise, distance from the interviewer, low speaking volume , or technical difficulties. For many quality improvement objectives, hearing both the respondent and the interviewer clearly and distinctly is critical. Thus, ascertaining CARI audio quality was perhaps the most important objective of the feasibility study.

To objectively rate the audio files, we developed a plan that included the development of an audio-quality rating system, the random selection of audio files from the NSCAW study, and file coding procedures that maximized agreement between audio quality raters.

The audio quality rating system consisted of four scale points - Excellent, Good, Fair, and Poor - and was developed from a perspective that focused on the envisioned uses of CARI:

- **Excellent Rating:** Both the interviewer and respondent are clearly intelligible. These audio files can be used for all foreseen purposes, including interview verification, assessment of interviewer performance, and questionnaire evaluation and pretesting.
- **Good Rating:** For these files, the rater can clearly hear the interviewer and can determine that a second party (presumably the respondent) is participating. However, some of the interchange between the two, particularly from the respondent, is unintelligible. Such audio files could be used for detection of interviewer falsification, but we may not be able to code the qualities of the interviewer-respondent interaction during the interview.
- **Fair Rating:** For these files, the interviewer can be heard and understood clearly all or most of the time; however, the respondent either cannot be heard at all, or can be heard but is mostly unintelligible. Because we are unable to hear a respondent's voice and to determine that voice is different from the interviewer, the file is not useful for verification purposes, and is of limited use for quality control purposes.
- **Poor Rating:** For these files, neither the interviewer nor the respondent is intelligible - the recording may simply consist of white noise or dead silence. These files are not useful for any envisioned purpose.

For assessing CARI audio quality, random samples of NSCAW caseworker and caregiver interviews were selected. The objective was to rate the quality of 100 files from caseworker interviews and 200 files from caregiver interviews. This sample was drawn with equal probability from the audio files produced by all NSCAW field interviewers. The 300 selected audio files were then divided among three designated raters.

Before starting the rating process, the three raters convened in order to ensure a high level of inter-rater reliability. During this meeting the raters listened to randomly selected segments of audio files. Raters indicated what they thought they heard, took notes, and rated the audio files. The raters discussed each audio file rating and established a consistent basis for rating the files.

A decision was made to listen to 10 verbal "exchanges" between the interviewer and the respondent for each audio file and assign a rating to this file using

the scale outlined above. If the file received a rating other than 'Excellent,' the coder also listed the reason why the file was not of the highest quality (e.g. background noise, technical problem, respondent voice quality). The results of this coding exercise are given in Table 1.

Table 1. Results of the Audio Quality Assessment

Rating	Caseworker		Caregiver	
	percent	*n*	percent	*n*
Excellent	94	94	92	184
Good	2	2	5	10
Fair	1	1	0	0
Poor	3	3	3	6

As seen from Table 1, over 90% of the files received the highest quality rating. For the six caseworker files that were rated less than "Excellent," the raters cited technical problems for five files and background noise for one file. For the 16 caregiver files that received a rating of less than "Excellent," the raters listed respondent voice quality for nine files and technical difficulties for seven files. These extremely positive results suggest that use of CARI technology does produce high quality audio recorded interviews. Confident of the quality of the recorded interviews, we could then turn to assessing other features of CARI feasibility.

Respondent Cooperation. A second step in determining CARI feasibility is to determine the extent to which respondents comply with the request to audio record their interviews using CARI. Based on an analysis of completed NSCAW interviews, we found that among cases that resulted in a completed interview, consent to use CARI was obtained in 85% of caseworkers interviews, 83% of the caregiver interviews, and 82% of the child interviews.

We view these consent rates as quite positive since they either meet or exceed the consent rates reported for conventional tape recorded field interviews (e.g., Fowler & Roman, 1992). These results are all the more positive when one considers the highly sensitive nature of the NSCAW survey (many parts of the questionnaire are self-administered) and the high emphasis given to obtaining signed and audio-recorded spoken consent to allow CARI, during the survey introduction process. Feedback we have obtained so far from both respondents and interviewers suggest that the very sensitive nature of the NSCAW interview causes some respondents to view the audio recordings with suspicion, possibly to be used against them in a court of law. Thus, the cooperation rates observed for the NSCAW may represent a worst case scenario for CAPI surveys. We expect to achieve somewhat higher compliance rates for field surveys that involve topics that are less specifically sensitive. Still, despite this limitation, CARI consent rates are still quite acceptable for all its envisioned uses.

We are also currently exploring ways to increase these cooperation rates by examining our informed consent statements and procedures, and by collecting information from interviewers in an effort to understand why some respondents are refusing CARI recording.

Interviewer Reactions to CARI. Another key criterion in the feasibility of CARI for production surveys is the reactions of data collection staff, and particularly interviewers, to the use of this technological innovation. The NSCAW CARI q.c. monitors report that the information content contained in the CARI recordings is excellent. In addition, the government project officer, the RTI project director, and data collection supervisors are all extremely positive towards CARI.

To assess interviewer reactions to CARI, we developed an interviewer debriefing questionnaire to gather their reactions and concerns about CARI. The questionnaire consisted of items about how respondents (caseworkers and caregivers) reacted to CARI, and also asked the interviewers to provide information about their perceptions of CARI from multiple perspectives. The interviewer debriefing questionnaire was sent to 65 interviewers who had significant experience in conducting NSCAW interviews and using CARI (at least 10 interviews). A total of 62 were received, for a 95.4% response rate (three interviewers terminated employment and did not return the questionnaire).

These results were also quite encouraging: 82% of the interviewers reported feeling extremely or somewhat positive or neutral about the overall use of CARI, compared to 18% who felt extremely or somewhat negative. Approximately 90% of the interviewers felt either extremely/somewhat positive, or neutral, about using CARI as a way to evaluate and provide feedback to interviewers. Further, 87% of interviewers felt extremely/somewhat positive about using CARI as a falsification detection method.

In reporting their feelings concerning respondent reactions, approximately 70% of the interviewers reported that respondents (caseworkers and caregivers) either "never" or "rarely (i.e., only a few times)" reacted negatively about the request to use CARI in the interview. Approximately 80% reported that after the interview commences, respondents seemed to be thinking about CARI either "not at all" or only "a little."

Some interviewers commented that they felt reassured in threatening interview situations knowing

that CARI was creating a audio record of the encounter. Others observed that CARI provides an opportunity for supervisors to appreciate the difficult tasks interviewers perform for the NSCAW and view CARI as a necessary part of interview quality control in a large field operation.

Additional data analysis suggested an inverse relationship between prior interviewing experience and acceptance of CARI: more experienced interviewers were more likely to exhibit negative feelings towards the use of CARI. The reason for this may be that some experienced interviewers view CARI as a departure from the status quo and resist CARI on the grounds that it is new. In addition, open-ended comments in the debriefing survey suggest that some interviewers view CARI as an sign that "management" does not trust the interviewers.

We feel that with the continued use of CARI on the NSCAW project, and hopefully an expansion to other projects, experienced interviewers will become accustomed to using CARI and will perceive CARI as a new standard survey work. When this occurs, we believe that the perception that CARI is an intrusion on the interviewer's privacy will dissipate. Further, we propose that the introduction of CARI recording mainly as a means to monitor interviewer performance in order to enable helpful feedback might increase the degree to which the procedure is well-received by all interviewers.

Costs of Implementing CARI in a Production Survey. Finally, we examined the costs of using CARI within a production survey environment for the limited purpose of interview falsification detection, and compared these costs to that of two traditional methods of verification. For the traditional approaches, we assumed that the basic mode of recontact to verify that the interview was conducted is the telephone. For households in the verification sample that are telephone nonrespondents or that cannot be reached by phone (non-telephone households), two alternative contact modes were considered: one using a mail post-card verification approach and another using a face to face contact verification approach. The former method is cost effective but will usually result in a very low rate of return (20 percent or less). The later has a higher contact rate but is much more expensive.

For a hypothetical survey, we applied the following cost assumptions for modeling the costs of the two telephone-based approaches and the CARI-based approach:

- the sample size for the hypothetical survey is 5,000 interviews,
- 10% of this sample is selected for verification purposes for all three approaches,
- the survey is conducted in 50 PSUs selected from the entire U.S.,
- the number of interviewers for the survey is 100 interviewers
- two interviewers work in each PSU (this affects the telephone-based approach involving face-to-face follow-up of nonrespondents and non-phone households)
- the phone coverage rate for target population is 90 percent (this affects both telephone-based approaches),
- the consent rate for allowing CARI recording of interviews is 85 percent (this affects the CARI-based approach only),
- 10 percent of the CARI verification sample is re-verified using the telephone-based approach involving face-to-face follow-up.

The costs for the hypothetical survey were based upon RTI's actual experience from the NSCAW as well as several other recent CAPI studies that have implemented the telephone-based verification schemes. In each case, we ignored system development costs; i.e., we assumed that a fully functional CARI-based or telephone-based verification scheme exists in the organization which requires little or no modification to accommodate our hypothetical survey.

Our analysis showed that the CARI-based verification approach was less expensive than either of the traditional approaches. Compared with a telephone/postcard scheme, the ratio of CARI costs per sample case to telephone costs was 0.68. That is, the cost of the CARI-based system 32 percent less than the telephone-based approach with postcard follow-up. Compared with the telephone/face to face approach, this ratio was 0.77; i.e., CARI cost 23 percent less than the telephone with a face-to-face follow-up approach. The finding that CARI saves money in survey project work suggests a further advantage of this system as currently configured.

3. Summary and Future Research

The objective of this research was to determine the feasibility of CARI for a wide range of survey quality improvement purposes, including the detection/deterrence of interview fabrication, interviewer performance monitoring, questionnaire evaluation and usability investigations, and collection of verbatim survey responses. Our results indicate that CARI is feasible for all of the above uses. It provides an unobtrusive means of recording the interactions between the respondent and the interviewer. CARI audio quality is comparable to analog recordings using tape recorders, but without the logistical problems that accompany the use of external audio recording devices in large-scale field operations.

Interviewers seem generally favorable or neutral towards the use of CARI to monitor their work. However, 10-15 percent of interviewers were mildly to very negative toward the use of CARI and viewed it as an intrusion on their privacy or a sign that management "doesn't trust us." More experience interviewers seem more resistant to CARI.

Although 82-85 percent of NSCAW respondents consented to the use of CARI, we expect consent rates to be somewhat higher in surveys with less sensitive content. Interviewers report no particular issues associated with introducing CARI to respondents, and suggest that respondents appear not to think about the recording during the interview. However, we are currently conducting a respondent debriefing survey for the NSCAW that will address this question more definitively. In particular, respondents are asked immediately following the interview about their reactions to CARI, and whether CARI has in any way influenced their responses.

The results of the cost analysis were also quite positive. A CARI based verification system can save 20 to 30 percent of the cost of traditional reinterview approaches to interview verification. However, this cost does not include the costs of using CARI for interviewer performance monitoring and feedback, which we believe should be part of the CARI q.c. system. Like call monitoring operations in telephone call centers, the costs of field interview monitoring will vary depending upon the frequency and intensity of the monitoring.

RTI and the Census Bureau are currently investigating cost-effective approaches for monitoring interviewer performance, including timely feedback to the interviewers to improve their performance. For new interviewers, CARI is being tested as a tool for extending interviewer training to the field environment: Newly trained interviewers are monitored intensively for several months after initial training, with provision for frequent feedback to provide mentoring, guidance, and suggestions for improving performance. This investigation assesses the reactions of interviewers and supervisors to this level of monitoring, and the effect of monitoring on interviewer performance and attrition.

Other CARI implementation issues that are currently being explored include:

- The use of CARI with decentralized telephone interviewing (from interviewers' homes) using plug-in telephone microphones.
- Options for encrypting the CARI audio files prior to transmission in order to protect confidentiality.
- Issues associated with the use of CARI in the context of the need for protection of respondent confidentiality.

References

Biemer, P. and Stokes, L (1989). "The Optimal Design of Quality Control Samples to Detect Interviewer Cheating," *Journal of Official Statistics*, 5(1), 23-40.

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

Fowler, F.J., & Roman, A.M. (1992). *A Study of Approaches to Survey Question Evaluation.* Final Working Paper submitted to the U.S. Bureau of the Census. Center for Survey Research, University of Massachusetts.

Massagli, M.P., & Clarridge, B.R. (1995) "Quality Improvement Through the Use of Systematic Monitoring in a Centralized CATI Facility," *Proceedings of the International Conference on Survey Measurement and Process Quality*, Bristol, United Kingdom. Alexandria, Virginia: American Statistical Association: 18-23.

Mudryk, W., Burgess, M.J., & Xiao, P. (1996) "Quality Control of CATI Operations in Statistics Canada," *Proceedings of the Section on Survey Research Methods*, JSM of the American Statistical Association, Chicago, IL. Alexandria, VA, 150-159.

Stanley, J.S. & Erth, S. (1995). "Development of a Field Monitoring Program for Telephone Interview and Instrument Quality Evaluation," *Proceedings of the International Conference on Survey Measurement and Process Quality*, Bristol, United Kingdom. Alexandria, Virginia, 24-28.