# Article

## Statistics Canada's Computer-Assisted Personal Interview Monitoring Program

by Caroline Pelletier and Edward Joseph

2009

Statistics Canada    Statistique Canada

Canada

# Statistics Canada's Computer-Assisted Personal Interview Monitoring Program

Caroline Pelletier and Edward Joseph[1]

## Abstract

Over the past year, Statistics Canada has been developing and testing a new way to monitor the performance of interviewers conducting computer-assisted personal interviews (CAPI). A formal process already exists for monitoring centralized telephone interviews. Monitors listen to telephone interviews as they take place to assess the interviewer's performance using pre-defined criteria and provide feedback to the interviewer on what was well done and what needs improvement. For the CAPI program, we have developed and are testing a pilot approach whereby interviews are digitally recorded and later a monitor listens to these recordings to assess the field interviewer's performance and provide feedback in order to help improve the quality of the data. In this paper, we will present an overview of the CAPI monitoring project at Statistics Canada by describing the CAPI monitoring methodology and the plans for implementation.

Key Words: Interviewer monitoring, Quality improvement.

## 1. Introduction

Computer-assisted personal interviewing (CAPI) operations require interviewers to have good interviewing skills. They must understand the subject matter concepts related to the survey well, maintain a good rapport with the respondent, and deliver the questions appropriately. All of these things must be done simultaneously in a production environment. Monitoring interviewers provides an opportunity to evaluate their skills and improve the interviewing process. Monitoring in the context of CAPI operations involves a third person, the monitor, who listens to portions of digitally recorded interviews and assesses the interaction between the interviewer and the respondent based on pre-identified criteria. Providing a statistical framework for monitoring allows quantitative and objective measures to be taken and effective feedback to be given to interviewers.

This paper describes the methodology and the plans for implementation of a CAPI monitoring program at Statistics Canada, using an audio method of monitoring.

## 2. Existing methods of monitoring at Statistics Canada

### 2.1 Computer-assisted telephone interview monitoring

Computer-assisted telephone interview (CATI) monitoring has been in place for several years (Mudryk et al., 1996) and is the baseline from which the CAPI monitoring project has been designed. The goal when developing the CAPI monitoring methodology was to make it as consistent as possible with the one developed for CATI monitoring. Call-centres are outfitted with a monitoring station located in a remote area where monitors (who are usually individuals supervising the interviewers) listen to live interviews using a telephone listening device while they see the interviewer's work being entered on a duplicate image of the computer screen. This screen shows the current question and the respondent's answer, as entered by the interviewer. The monitors evaluate the quality of the interviewer's work against a pre-defined set of criteria and notes on a form when an error is made or poor

---

[1]Caroline Pelletier, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (pellcar@statcan.gc.ca); Edward Joseph, Statistics Canada, 170 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (joseedw@statcan.gc.ca)

interviewing practices are used. This information is then used to provide summary statistics and feedback to interviewers, supervisors as well as managers. The summary statistics include error rates and listings of errors. Immediate feedback is provided for critical errors; for all other types of errors, feedback is done on a weekly basis. It is done through personal and group discussions using the information available in the reports. The frequency at which an interviewer is monitored is based on previous performance. For example, an experienced interviewer who has consistently met or exceeded performance expectations will not be subject to the same amount of monitoring as a new Statistics Canada interviewer.

## 2.2 Verification

Verification of personal and telephone interviews is another method of monitoring used at Statistics Canada, where call-backs are made to respondents in order to ensure that an interview actually took place with them. Monitors ask the respondents whether they were contacted by a Statistics Canada interviewer and if they participated in the particular survey. They then ask some key survey-specific questions to make sure the complete interview took place. Though this method can be effective, a significant burden is placed on the respondent by re-contacting and asking survey-specific questions a second time. Moreover, the respondent may not recall the specific answers provided previously. This method is also costly. For these reasons it is rarely used anymore at Statistics Canada.

## 3. Objectives of the CAPI monitoring program

The CAPI operations involve the Statistics Canada interviewer using a laptop computer to key in responses to an electronic survey questionnaire during a face-to-face interview. The interview is normally conducted in the respondent's home or, although rarely, in another agreed upon location. Once the interview is completed, the data are transmitted overnight to Statistics Canada Headquarters for processing. No measure of field interviewer performance is currently available.

There are many objectives to the CAPI monitoring project. It is meant to provide an objective and consistent measure of field interviewer performance over time. It also ensures coherence of collection within and between surveys. It allows the monitors to ensure that survey-specific procedures are followed in the field as it relates to data collection as well as ensuring that basic interviewer procedures are being followed. The monitoring project also aims to improve data quality through timely and effective feedback to interviewers, supervisors, and data collection managers. CAPI monitoring will allow for correction or support of interviewer performance in addition to detecting or deterring falsification of interviews.

## 4. Overview of the CAPI monitoring methodology

The CAPI monitoring strategy involves the audio recording of pre-identified snippets of interviews. The information is saved in audio files which are transmitted daily from the field to the headquarters. Monitoring takes place in the regional offices using the information available on the headquarters' servers. Based upon a sampling plan, the monitors listen to a certain number of audio files per interviewer on a weekly basis while following the flow of the interview on a paper questionnaire and assess the interaction between the interviewer and the respondent. Unlike CATI monitoring, the monitor does not have a duplicate image of the interviewer's computer screen so it is not possible to monitor a characteristic such as the accuracy of the data capture. In addition, non-verbal communications that take place but are not recorded cannot be monitored. During a monitoring session, if an interviewer makes an error or shows poor interviewing practices for any question, this information is recorded by the monitor on a form and later tabulated. Reports on the quality of the interviewing process are available for the interviewers, the supervisors, and management. Timely feedback is then provided to the interviewers for critical errors and weekly feedback is provided for other types of errors. Unlike CATI monitoring, it is not possible to provide real-time feedback for critical errors. Although audio recording can be used to directly collect respondent data or to evaluate and improve a questionnaire through behavioural coding, the purpose of this CAPI monitoring program is restricted to evaluating interviewer performance and providing feedback.

# 5. Privacy issues

The recording contains confidential information that must be protected. A Privacy Impact Assessment (PIA) had to be conducted for the CAPI monitoring program. This is an evaluation process that requires those responsible for the collection, use, and disclosure of personal information to evaluate the privacy, confidentiality, and security risks that may be involved in a project and to develop mitigation measures aimed at avoiding or reducing the identified risks.

Hence, the objectives of the privacy impact assessment are to:

- ensure privacy protection is a key consideration in the framing of this project;
- describe the data flow of all transmissions from the interviewers;
- identify any potentials risks and impacts to the confidentiality of respondent data as well as any privacy issues related to interviewers;
- outline the safeguards employed to address the potential risks.

This assessment of the CAPI monitoring program did not identify any risks that cannot be managed using existing safeguards for the protection of confidential survey data and putting in place additional safeguards specific to this program.

In the majority of cases, the recording will be kept for a maximum of three months before being destroyed. In limited cases, it may be necessary to keep the recordings for a longer period of time (up to a maximum of a year). In the latter type of cases, the exact retention period will be decided at the time that a request for extension to the retention period is made. The manager of the database will have the ongoing responsibility to ensure that recordings are destroyed as per the established retention periods.

To comply with Statistics Canada's strict security standards for the protection of data, the transmission of all files containing recordings of the interview follows the same data flow and security requirements as survey data. However, the recordings of interviews are separate audio files that are transmitted directly to Statistics Canada's Collection Planning and Management Division while survey data go to the subject-matter division responsible for the survey.

# 6. Monitoring characteristics

A major decision in developing a monitoring scheme involves identifying what behaviours should be monitored and exactly how they should be measured. The characteristics used to evaluate the CAPI process are based upon desirable and undesirable behaviours in the interviewing aspect of the process. Having set criteria allows consistency in the evaluation of one CAPI interviewer to the next. All interviewing characteristics used in CATI monitoring were reviewed and most of them were kept to ensure consistency in the evaluation of CATI and CAPI interviewers. These characteristics relate to the main interviewing functions: question delivery, respondent relations, and subject-matter knowledge and are presented in Table 6-1. Characteristics pertaining to data entry and processing, which are considered in CATI monitoring, are currently excluded because monitors only have access to the audio file and are not able to see what was entered by the interviewer. No additional characteristics were added as part of the CAPI monitoring project. Different weights are assigned to errors based upon their seriousness. This will be discussed in detail in section 9.

**Table 6-1**
**Error category by interviewing function**

| Interviewing Function | Error Category |
|---|---|
| Question delivery | Asking |
| | Wording |
| | Professionalism / Voice |
| Respondent relations | Probing / anticipation |
| | Judgement |
| Subject matter | Definitions |
| | Concepts |

# 7. Monitoring methodology

## 7.1 Sampling unit

The sampling unit in quality control is usually the unit of product or related process over which we want to establish some form of control. For CAPI operations, we want to monitor the interviewing process, more specifically by individual interview. Only parts of the interview are recorded because recording a complete interview is not a viable option due to the fact that a large file would need to be transmitted from the field to the headquarters via a phone line. Recording parts of the interview is possible through the use of pre-identified trigger questions which start the recording. It is not possible to start recording parts of the interview at a random point. The recording can end at another pre-identified question, at the end of the block of questions or after a certain amount of time. Trigger questions must be identified during development and cannot be changed in production. Each part of the interview that is recorded is called a snippet. Recording will not take place if the trigger question is skipped as part of the questionnaire flow. For this reason, more snippets are recorded than might be needed.

When recording parts of the interview, it is possible to record blocks of questions or turn the recording on for a certain amount of time. In the context of CATI operations, the sampling unit is a screen which usually corresponds to a question. For CAPI monitoring, blocks of questions instead of timed recordings are recorded to make the methodology similar to the one used for CATI monitoring. It also facilitates the monitor's work because the person is able to follow easily the flow of the interview on a paper questionnaire and does not have to juggle between sections of the questionnaire. Each block of questions recorded consists of 5 to 7 questions. A question is considered as the sampling unit. The recording either ends at another pre-identified question or at the end of the block of questions.

For CATI monitoring, a sample is comprised of 20 screens or questions. In CAPI, 20 questions within a number of blocks of recorded questions form a sample. Recording 5 to 6 blocks of 5 to 7 questions per interview is sufficient to provide a large enough sample on which interviewer performance can be evaluated. It also allows a sufficiently large numbers of characteristics to be monitored.

## 7.2 Sampling plans

All interviewers do not need to be sampled at the same rate over time. Some interviewers have very good interviewing skills and do not require the same amount of monitoring as new Statistics Canada interviewers. As is the case with CATI monitoring, interviewers are stratified into 4 groups based on their previous performance:

- (A) Experienced and excellent interviewers
- (B) Very good interviewers
- (C) Acceptable or new interviewers
- (R) Unacceptable performers

When monitoring starts for a given survey, all interviewers, including new Statistics Canada interviewers, are placed into group C. Over time, as interviewers improve their skills and abilities, they are moved from group C to B and likewise, from group B to A. Those who perform at an unacceptable level in comparison with the standards set for their plan are demoted to the next lower level (i.e. from group A to B, B to C, and C to R) so that more of their work will be monitored until their performance improves. The promotion and demotion approach is managed automatically. This approach is similar to that used in CATI monitoring.

## 7.3 Sample size

Once in production, the number of samples selected for CAPI monitoring will depend on the interviewer's group. One sample per week will be selected for interviewers in group A, 2 in group B, 3 in group C, and 4 in group R. The amount of monitoring resources available, a factor to be considered when determining the sample size, is unknown at this point so the proposed numbers may need to be adjusted.

## 7.4 Sample selection

Periodic samples must be selected to monitor the interviewing process at specific points in time. For CAPI operations, this is achieved as follows:

- Sort the audio files collected during a work week in sequential order using the date and time stamp associated with each file.
- Use systematic sampling to select the sample. Knowing the number of samples to be selected for an interviewer (n), separate the population of audio files collected during the week (N) into intervals by dividing it by the number of samples to be selected. This is known as the sampling interval (k). Randomly select a number between 1 and the end of the sampling interval (k). This is the start point (r). The first recording to listen to for each sample is respectively the $r^{th}$, $(r+k)^{th}$, …, $[r+(n-1)k]^{th}$ files. By using systematic sampling, this ensure that the work of each interviewer is sampled at approximately fixed intervals of time within any given work week.
- For each sample, monitor the next 20 questions for the interviewer and record errors for each quality characteristics that did not meet the requirements. These questions could come from one or more consecutive audio files and one or more interviews.

# 8. Field collection

External microphones are used to record portions of each interview as they provide the best audio quality. A recording application integrated within the collection application (Blaise) is used to automatically turn the recording on and off at the appropriate questions as well as at the consent question to record question and answer. It must be possible to turn the recording on and off without a detectable effect on performance and being noticed by interviewers. The respondent is asked the consent question for approval for recording, after having given approval to participate in the survey. It must be asked each time the respondent is contacted. If consent is not provided, the recording is disabled within the application. In addition, it must be possible to stop the recording at any point during the interview if the respondent requests it. In this situation, the interviewer must close the CAPI application to terminate the recording and then re-start it to continue the interview. Skips due to certain responses or refusals are unavoidable and affect the length of recording and the capability of the instrument to turn the recording off at the appropriate time. Therefore, extra logic must be built into the application.

The blocks of questions to be recorded are selected throughout the questionnaire including the entry and exit modules which contain introduction, specific instructions, and closing statements. The questions in the blocks chosen for the recording must be asked to most respondents. In order to obtain a good representation of interviewer skills, easy and difficult questions that are somewhat representative of the questionnaire as a whole are recorded. Difficult questions can use complex terminology that may require the interviewer to provide more information to the respondent without leading the person towards a certain answer or cover a sensitive topic such as physical abuse

or gambling problem. Each block of questions is saved in an audio file using the native Windows software (in .wav format). The audio file is then compressed to reduce its size using an open source software called Speex and transmitted to the headquarters.

In order to transmit the data, interviewers dial in with authentication over public lines using the laptop computer and upload the encrypted questionnaire data files to an application-specific server at Statistics Canada Headquarters. Via specialized switching technology, the encrypted files are transferred from the publicly accessible network to Statistics Canada's secure, private network for further processing. Files are decrypted only after they have been transferred to this network. The use of phone lines to transmit large audio files is problematic as these files may slow down the system or create bottlenecks. The use of high speed internet may be the solution to this problem in the future.

## 9. Classification of errors

Some error situations are considered to be more serious and should be assigned higher weights than others. Thus, using coarse language has more impact than reading too slow. The same demerit rating system as for CATI operations is used for CAPI:

> Critical error = 4 demerits
> Major error   = 2 demerits
> Minor error   = 1 demerit

The same weight always applies for a certain error. For example, the use of coarse language is always a critical error regardless of the severity. The weights are not based on the monitor's judgement, are consistent with those used in CATI monitoring, and do not change from one survey to the next thus allowing consistency in the evaluation of interviewers. By applying this demerit rating system, the accumulation of minor and/or major errors can be equivalent to at least one critical error.

For CAPI operations, the quality measure that relates to the interviewing ability of the interviewer is the total number of defects encountered within a sample of 20 questions being taken over a number of audio files recorded. The same error can then occur more than once in a sample since each question can have the same problem repeated. For example, the interviewer may not read the question as worded in two instances. The cumulative effect of all errors is used in determining the overall quality of the interviewing process. By applying weights to the errors, weighted quality measures (total weighted error count per 20 questions) are computed for each interviewer and summary statistics are provided for the overall CAPI operation.

## 10. Feedback

During a monitoring session, errors and poor interviewing practices are recorded, as they are discovered, on a paper form by category and later entered into a database. The errors are tabulated weekly at the interviewer level and for all interviewers. Reports containing different levels of information are available for interviewers, supervisors, and management. This allows problems to be quickly identified. In order to improve the quality of the interviewing process, timely feedback of results must be provided to interviewers. Corrective feedback is provided to prevent problems from reoccurring and positive feedback is also provided to help maintain behaviours that have a positive impact on their performance.

Very timely feedback must be provided for critical errors. Unlike CATI monitoring, immediate feedback is impossible because monitoring can take place the day after the interview at the earliest. This poses a serious problem in the context of CAPI operations. Therefore, if the monitor records a critical error, he/she sets up an appointment with the interviewer immediately after the monitoring session. For major and minor errors, weekly feedback is provided.

## 11.  Pilot test

Before the CAPI monitoring program described in this paper is put into production, a pilot test will be conducted in November and December 2008 with the Canadian Community Health Survey. Three sites representing a mix of English and French interviews and of rural and urban areas were selected across the country. The consent question and answer will be recorded. Based on a study done at RTI (Arceneaux, 2007), it is expected that about 85% of the survey respondents in these sites will agree to have their interview recorded. This would represent 265 interviews recorded. Approximately 25 interviewers will conduct these interviews. In addition to the consent question, 7 blocks of questions considered to be representative of the questionnaire will be recorded to allow a reasonable evaluation of the interviewer's performance. Debriefing questionnaires will be administered to interviewers and respondents, regardless of whether they agreed to be recorded.

The pilot test will help us answer the following questions:
- What is the quality of the recording?
- Are there differences in the survey refusal rate due to recording?
- What is the cooperation rate for recording?
- Is the monitoring methodology feasible in real-life?
- What are the reactions of the interviewers and the respondents to being recorded?

The results from the pilot test will be analysed and recommendations will be made before moving to a production mode. This could involve modifications to the methodology described in this paper. The telephone line problem for data transmission will also need to be resolved by possibly using high speed internet. In addition, an automated sampling system will be developed, all interviewers and monitors will be trained, and the reports required for each group (interviewers, supervisors, and managers) will be identified. More automated tools will be developed to assist the monitors in carrying out the monitoring.

## 12.  Future work

Once the CAPI monitoring project goes into production, the framework could be used in other ways to improve the quality of the collected data. Measures of keying and coding errors could be produced if monitoring was done using an audio-visual method, meaning that the monitor could see what was entered by the interviewer. This would require that the Blaise audit trail file (a file of keystrokes) be integrated with the monitoring application in order to identify if the interviewer selected the right response category or typed a wrong answer.

The current project focuses on using audio recording to monitor interviewer performance. However, the audio recording technology can be used in many other ways to promote better quality of questionnaires and data. Subject-matter areas could use recordings to assess the content of the questionnaires, to identify potential problems with question wording or understanding by the respondents. Recording can also be used to directly capture responses to questions for subsequent coding or to assess how well these responses are coded into pre-defined categories by the interviewer.

## 13.  Acknowledgements

# References

Arceneaux, T. (2007). Evaluating the Computer Audio-Recorded Interviewing (CARI) Household Wellness Study (HWS) Field Test, *Proceedings of the Survey Research Methods Section*, American Statistical Association, 2811-2818.

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