

# Listening for Data: New-Millennium Prospects for Computer-Assisted Interviewing in Statistical Surveys

James E. Smith, Ph.D.  
Westat  
1650 Research Blvd.  
Rockville, MD 20850 USA

## 1. Computer-assisted Interviewing (CAI) Today

Today it is common for a field interviewer to arrive on a doorstep carrying a laptop computer, and perhaps a hand-held computer too. Even more common are interviewers wearing telephone headsets and reading questions from a computer screen while tapping responses on a keyboard. Less common, but technically very feasible, are respondents sitting at their Internet browsers answering questionnaires that have text, graphics, and sound.

There are a many reasons why paper-and-pencil surveys are being replaced by various forms of computer-assisted personal (CAPI), telephone (CATI), and self (CASI) interviewing. Although the specific reasons vary from survey to survey, some common themes are : to increase data quality by using such features as computer-controlled question flow and automated data checking during interviewing; to improve timeliness of survey results by reducing the need for post-collection processes like key entry, manual data cleaning, and missing data retrieval; to make it possible to conduct increasingly complex surveys that require larger and more sophisticated questionnaires than could be reliably administered without computer assistance; and to reduce total survey costs relative to the quantity, quality, and timeliness of the data being collected.

Overall, practical experience with the various modes of computer-assisted interviewing (CAI) has been positive so far. As a result, today there are fewer issues concerning whether to use CAI than about how maximise the benefits of CAI (Nicholls et al, 1997). This involves a range of technical, scheduling, and cost considerations, all of which must be addressed in the context of rapidly advancing CAI technologies and innovations.

My purpose is to attempt to anticipate some major prospects for CAI in survey research based on three reflections about CAI experience to-date. These reflections are: (a) that CAI means much more than just computerising questionnaires, (b) that CAI innovation relies heavily on technology markets and trends, and (c) that many statistical surveys involving CAI have extended far beyond questionnaire data collection into becoming more comprehensive statistical data systems. Based on these reflections, I will conclude with some suggestions on prospects for CAI in the new millenium of survey research which is upon us.

## 2. CAI as More Than Computerising Questionnaires

Questionnaires are the *sine qua non* of survey research and CAI began with the goal of computerising hardcopy questionnaires. But for several reasons, today CAI means much more than programming a hardcopy questionnaire into a computer. One reason is that screen layouts, question wording, question flow, and computations in a computer questionnaire usually differ significantly from what is or can be done on paper. Another reason is that too close an imitation of paper may deny many of the benefits the computer can provide, an obvious example being dynamic fills and automated data checking. However, these additional benefits of a computerised questionnaire are only achieved with more, and more complex, up-front specification,

development, and testing CAI questionnaires than was required for hardcopy, thus changing the process of designing and developing a survey.

In addition to computerised questionnaires, CAI surveys almost always require some sort of automated case management system along with an electronic data communications system. These computerised components must also be specified, set up or developed, operationally integrated, and tested before being placed into use. This adds further to up-front survey design and development processes, but again with expected benefits in response rates, operational efficiency, and total survey cost and timeliness.

Thus, in a narrow sense CAI may mean computerising a hardcopy questionnaire, but in its more in a broader sense CAI is about new and enhanced kinds of questionnaires made possible by the computer along with computer enhancements to other critical survey processes. This broader meaning of CAI is evident in some interchangeable terms, such as the American term "CASIC" (computer-assisted survey information collection) or the European term "CADAC" (computer-assisted data collection) (de Leeuw and Collins, 1997; Couper and Nicholls, 1998). Recent symposium volumes present many discussions about CAI in this broader sense (Couper et al, 1998; Lyberg et al, 1997).

### **3. CAI as Technology Adoption**

Because as survey researchers we are adopters, not creators, of the computer technologies we use, we rely upon the marketplace to provide useful and affordable technology solutions. However, survey research is not a large enough market to be a significant driver of technology trends, so it is up to survey researchers to adopt and adapt technologies for their needs. Thus, CAI innovations usually take the form of creative use of technologies, often with compromises between what is most desirable and what is possible or feasible given available technologies.

Some of the major computer technology trends that have been of particular relevance to survey researchers are:

- *smaller devices*, such as minicomputers, then PC's, followed by laptop and notebook PC's, and now hand-held computers;
- *expanded connectivity* starting with networks and modems, and now expanded by the Internet and wireless communications;
- *integrated multimedia* in which software and hardware regularly incorporate capabilities for graphics, sound, video, and various forms of tactile user interfaces;
- *component-based software* in which powerful computer capabilities are made available to programmers as highly packaged reusable software objects or components.

In attempting to capture the benefits of these technology trends for survey research, CAI developers have reached into the market for a mix of mainstream, emerging, and experimental technologies. Some important examples are:

*Mainstream technologies (major market presence; good technical stability and support; a few dominant products and vendors; relatively favourable pricing due to large market), such as:* the use of minicomputers, servers, and PC networks for CATI; laptop computers and modem communications for CAPI; cellular phones for dial-in CATI from non-telephone households; the use of commercial off-the-shelf software (COTS) packages where available for survey applications development.

*Emerging technologies (good market presence; perhaps features or performance limitations; but improving rapidly; survival of particular brands and vendors somewhat uncertain; often premium pricing due to limited market but may offer favourable pricing to gain market share), such as:* hand-held computers or convertible laptop computers for doorstep screening; the use of touch-screen or pen-based technologies (e.g. Bosley et al, 1998); audio-CASI (e.g. Tourangeau and Smith, 1998); integrated voice response (IVR) in CATI; many of the Web technologies for data collection or survey operations support.

*Experimental technologies (niche or limited market presence; products strong on concept but uncertain as to market viability; technologies may be superseded by yet unknown developments or may become incorporated as features into more established technologies), such as:* expert systems for data coding (e.g. Conrad, 1997); automated speech recognition for data collection (e.g. Blyth, 1997).

#### **4. CAI as Statistical Data Systems**

Regardless of the technologies they use, statistical surveys are ultimately successful when they produce useful data. In fact, a striking feature of today's largest and most complex CAI surveys is the degree to which their data collection extends well beyond questionnaires. Moreover, the questionnaires that are used emphasise gathering "factual" or "external" data more than attitudes or opinions, thus making the questionnaires very complex. In addition, these studies often require multiple data collections over time and multiple collections from different sources. The following are a few examples of current studies of this kind sponsored by agencies of the U.S. Government:

- a multi-round, multi-panel study of medical expenditures involving extensive household CAPI interviews and extensive collection of records data from medical providers;
- a health and nutrition survey that collects data from CAPI interviews, medical examinations, audio-CASI, and laboratory results;
- a longitudinal study of child development involving CAPI-like classroom testing and CATI interviewing of parents;
- a medical disability survey requiring large-scale CATI screening followed by second-stage CAPI screening, and then CAPI interviewing, medical examinations, and laboratory data collection.

These sophisticated CAI studies are "multi-mode" in the sense of using multiple forms of CAI (e.g. CATI, CAPI, audio-CASI, etc.) as well as other computer-assisted methods for various types of non-questionnaire data. In such studies, a critical success factor is the ability to integrate various CAI technologies and processes into an operational whole. In addition, the diversity of data sources requires an overarching "total survey statistical database" design which incorporates the various questionnaire and non-questionnaire data as subsets.

#### **5. Prospects**

As an applied discipline, survey research evolves largely through developments arising out of the practice of doing surveys. Taken individually and collectively, these experiences give concrete meanings to the somewhat abstract ideas of "survey needs" and "CAI capabilities." The outcomes of future survey experiences are unpredictable, but based on the reflections about current experience discussed above, some short term prospects for CAI arise:

Prospect #1: CAI questionnaires and other CAI survey components and processes will include more and more uniquely computer-enabled benefits. Traditional survey design and

development processes for specifying, communicating, and testing CAI questionnaires and other computerised survey components will not suffice to fully achieve these benefits and will need to be replaced before they become a bottleneck.

Prospect #2: Greatly expanded connectivity combined with highly integrated multimedia in virtually all computer technologies will force a fundamental rethinking of the purposes and methods of survey interviewing, including rethinking the roles of interviewers.

Prospect #3: Increasingly surveys will become statistical data systems that collect data at multiple times from multiple sources using multiple modes of CAI. The scale and complexity of these data will necessitate that “statistical database,” rather than “questionnaire,” becomes the central organising concept, with all forms and instances of questionnaire and non-questionnaire data as subsets..

These prospects converge to suggest that the foundations of twentieth-century survey research is evolving from its origins in “asking questions” (Payne, 1951; Sudman and Bradburn, 1987) into a new 21<sup>st</sup>-century foundation of “listening for data” in the form of highly computer-enabled, multi-media infused, multi-source, and often ongoing statistical surveys.

## REFERENCES

- Blyth, B. (1997). “Developing a Speech Recognition Application for Survey Research,” in Lynberg et al (1997), pp. 249-266.
- Bosley, J., Conrad, F. and Uglow, D. (1998). “Pen CASIC: Design and Usability,” in Couper et al (1998), pp. 521-542.
- Conrad, F. (1997). “Using Expert Systems to Model and Improve Survey Classification Processes,” in Lyberg et al (1997), pp. 393-414.
- Couper, M. et al (1998). *Computer Assisted Survey Information Collection*. New York: John Wiley & Sons.
- Couper, M. and Nicholls II, W.L. (1998). “The History and Development of Computer Assisted Survey Information Collection Methods,” in Couper et al (1998), pp. 1-22.
- de Leeuw, E. and Collins, M. (1997). “Data Collection Methods and Survey Quality: An Overview,” in Lyberg et al (1997), pp. 199-220.
- Lyberg, L. et al, (1997). *Survey Measurement and Process Quality*. New York: John Wiley & Sons.
- Nicholls II, W., Baker, R., and Martin, J. (1997). “The Effect of New Data Collection Technologies on Survey Data Quality,” in Lyberg et al (1997), pp. 221-248.
- Payne, S. L. (1951). *The Art of Asking Questions*. Princeton: Princeton University Press.
- Sudman, S. and Bradburn, N. M. (1987). *Asking Questions*. San Francisco: Jossey-Bass Publishers.
- Tourangeau, R. and Smith T.W. (1998). “Collecting Sensitive Information with Different Modes of Data Collection,” in Couper et al (1998), pp. 431-454.

## SUMMARY

This paper attempts to anticipate some of the major prospects for computer-assisted interviewing (CAI) in survey research as we enter the 21<sup>st</sup> century. Three reflections on experience with CAI are presented: (a) that CAI means much more than just computerising questionnaires, (b) that CAI innovation relies heavily on technology markets and trends, and (c) that many statistical surveys involving CAI have extended far beyond questionnaire data collection into becoming more comprehensive statistical data systems. Based on these reflections it is suggested that the foundations of twentieth century survey research in “art of asking questions” is evolving, and will to rapidly evolve into a new foundation of highly computer-enabled, multi-media infused, ongoing and extensive “listening for data.”