Implications of CATI

Costs, Errors, and Organization of Telephone Survey Research

ROBERT M. GROVES
University of Michigan

Computer-assisted telephone interviewing might best be viewed as an intervening technology. It is a vehicle through which survey researchers can do their work, and in that regard, it is similar to the relationship between the telephone system in general and telephone surveys. The telephone system was designed as a communication medium for friends and business acquaintances, and survey researchers are using it for a purpose for which it was not specifically designed. Whenever we attempt to push the medium—the telephone system—further than it was designed, we start suffering. So when we attempt to cover the entire population using telephone surveys, we pay a price because not all people have telephones; and we suffer if we want to introduce a group interviewing format, because the telephone is basically a one-on-one medium; and so, too, in CATI, we are using a technology to assist us in our work, and, as with the telephone system, it is a technology with already well-established features. The technology that we are adapting to the needs of survey research when CATI is used consists of computers for the storage and manipulation of questionnaire material and survey data, and cathode ray terminals (CRTs) for the presentation of questions and the input of responses. Relative to the use of the telephone system for surveys, this tool may be more adaptable to our needs because we can exert some control over what the technology is doing for us. We can, through programming alterntive features, set up different CATI characteristics that suit our particular needs. This power will always be limited, however, by the constraints of presenting question material to the interviewer on a CRT screen (that is, a single screen contents presented in discrete characters).

In the implementation of most new technologies, the first phase experienced is the use of the technology to solve old problems; most of the discussion of CATI in the past has revolved around issues arising

SOCIOLOGICAL METHODS & RESEARCH, Vol. 12 No. 2, November 1983 199-215

from such use. A secondary phase is the attention to new problems—some created by the technology, some issues that we can address for the first time. This article hopefully will point out areas given little attention in the past and areas that identify new problems the technology is bringing us. The remarks are divided into two main categories: (1) characteristics of hardward, software, and personnel organization regarding CATI; and (2) implications of CATI for the cost and error structure of survey data.

HARDWARE, SOFTWARE, AND THE ORGANZATION OF CATI

The paramount issue in making decisions about CATI now is to choose a point in the development of hardware technology that offers the most enduring efficiency. There is a constantly changing technological status and the movement clearly points to a growing dispersal of computing power. There are smaller and smaller units being created that are more and more powerful at a cheaper and cheaper cost. This question simplifies into one about how intelligent the terminals that interviewers use should be. Another issue is how large a system is actually needed, how much variation in workload needs to be accounted for. Most surrvey organizations have a certain base of production that needs to be done every month, but also unanticipated peaks and troughs of work. Unfortunately, hardware characteristics have to be chosen to provide for the peaks of demand. A corollary issue that is important for small organizations is the sort of multiple uses the computing resources offer. Some organizations are large enough that one could imagine a machine dedicated to on-line interviewing, perhaps coding also, but the analytic work could be done on other machines. Other organizations might best use the machine for CATI also for general analytic work, word processing, accounting, and so forth.

One orientation to the issue of software is common to a survey statistician who would examine cost versus variance models or cost versus error models, who is interested in viewing the financial savings—the fiscal changes that CATI might bring about—and also attempts to measure and estimate the changes in the error structure of survey data resulting from the CATI application. One of the major issues in software design is the cost efficiency of the production CATI system. The overriding issue that flows throughout that discussion is the decision regarding how much the interviewers should be doing, how many

decisions we allow the interviewers to have at each point, and how much the machine is doing for the researcher, the interviewers, and so on.

Finally, in deciding how much one invests in software development, we also have to make some estimate of how long a single system is expected to last before a rewrite is desirable. Some believe that systems will be in rewrite constantly; some think that a system will last maybe 3 or 4 years and then one would do a rewrite, either for reasons of changes in hardware, new software ideas, or new kinds of surveys that are being done.

One of the most fascinating problems that we are in fact just discovering is how CATI systems should be organized and administered within a survey organization. How should we inject this new technology into an existing survey unit? The smaller the unit, probably, the smaller this problem is, because single people can fill multiple roles. In a fairly large unit, however, where there are sections or departments devoted to particular phases in the survey act, two problems arise: (1) In what unit should responsibilities for the machine the placed? and (2) How can the existing units best make the transition to procedures that make them interdependent through the procedures of the CATI system?

Figure 1 attempts to describe what happens in telephone surveys in a non-CATI environment. The dimension that moves from left to right is time; the rows of the figure are units in the organization. In many organizations, there are separate staffs for sampling, field, and coding. Sometimes there is a separate staff for data processing; sometimes that is part of the research staff. The boxes on that page are points in the flow of work where those discrete units have to exchange information or materials. Since organizations have lived with this flow of work for years, the sampling section, for example, knows that it is linked to the research staff at the beginning to describe the sample in full, and then somewhat later they have to pass sampling materials to the field department, so the field department can begin their work. These are very narrow and well defined relationships between the different groups.

Figure 2 shows the effect on that organization when we introduce a machine that does on-line interviewing, represented by the rectangle in the figure. The machine is a piece of hardware, first of all, but it contains programs of various sorts. The user purchases an operating system from the manufacturer or acquires it in some way. There is programming development for the CATI system that is the left portion; there is application code for a particular survey, and then there is a set of files that are created by the survey. What has been done to that discrete

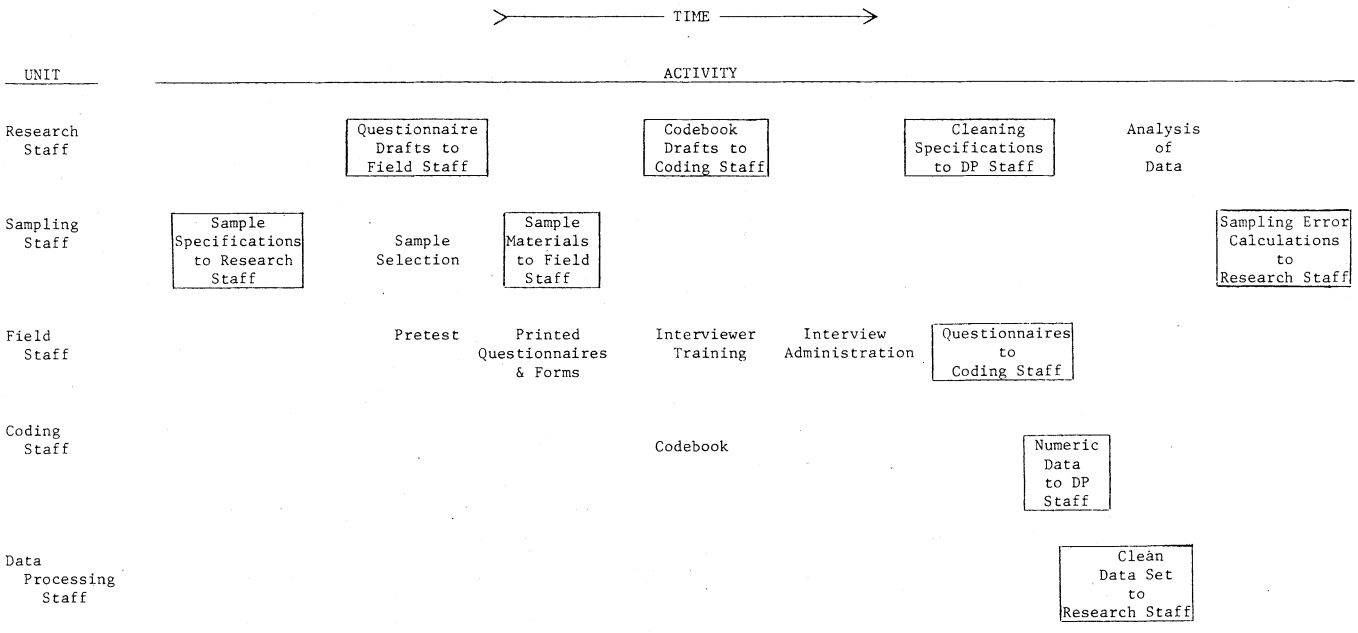

Figure 1: Telephone Surveys Under Non-CATI Procedures

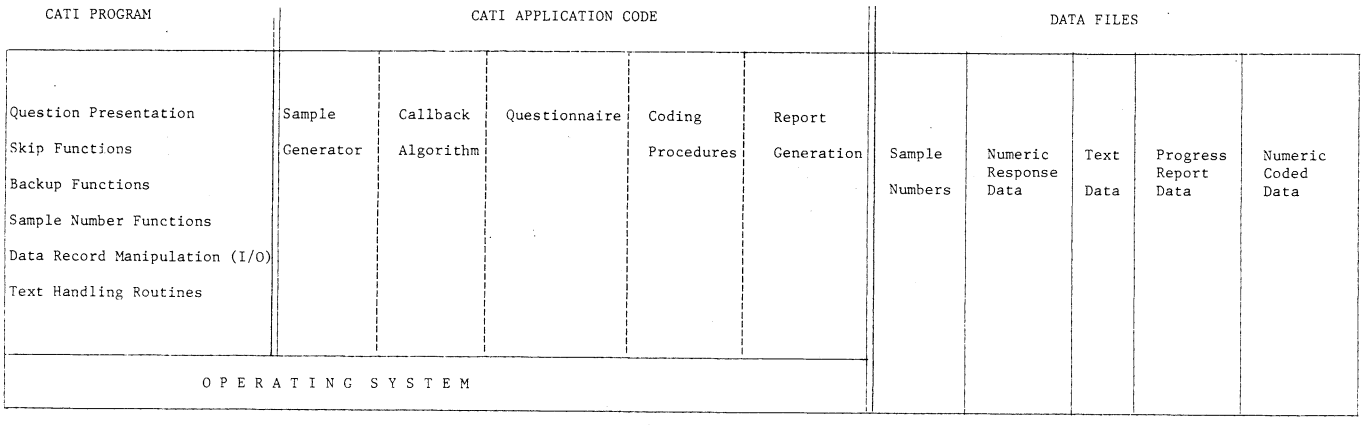

Figure 2: Telephone Surveys Under CATI Procedures

structure previously used is to place, at the center, a machine and software procedures that demand certain relationships among sections. The dotted lines between the different components of the CATI application code define the new relationships between these sections—sampling, field, and coding. All of the work is now designed to prepare the needed components for a survey on this one machine. The final deadlines are about the same, but the order of some tasks, if altered by the needs of the CATI system and the different work groups, must share resources more directly.

Now everyone is dependent upon the same resource; they are all using a single machine. The degree of coordination between the different sections is dependent on exactly what solution is taken to defining the sampling algorithm, when that has to be ready with respect to the questionnaire design, when and how consistency checks are to be placed on the system, and so on; but all groups are using the same terminals and they have to coordinate their design in ways that they did not have to before. In addition, the order of completion of tasks has been altered somewhat. The research staff in the old model could delay sending cleaning specifications to data processing until after the field work had begun, but in CATI, that must be part of the questionnaire setup. Instead of the iterative cleaning operations typical of post-coding processing, under CATI, all the cleaning operations can be accomplished during data collection. To achieve this, however, the researcher must specify all consistency checks before the interviewing begins.

The variability across CATI systems regarding this figure is related to how dotted are the lines, how firm are the boundaries, for example, and can sampling do their sample generation routines independently of the questionnaire development? How dependent is the questionnaire translation system, or is running pretesting of the questionnaire dependent on the sample design work, and so on? The questions on the callback algorithm and the report generation algorithm are really ones of—how transient are those? How often do they change? Are they fixed structures? Should they be flexible? Should we be able to change them on every survey?

Organizations are just now experiencing what will happen when this new technology is introduced in a production mode to personnel who have been accustomed to interacting in a different way. The transition to that state will, perhaps be aided by the fact that, in many organizations, these groups are now talking to one another more and more in terms of shared computer files, because machine assistance is entering already in

various ways. But the sharing of a machine directly and the complete coordination of their work is something new.

POSSIBLE CHANGES IN SURVEY COSTS AND ERRORS WITH CATI

Tables 1-4 present a listing of several components of a CATI system that cause changes in the design or implementation of a survey. Some of these changes may offer reduced survey costs or errors to the researcher, some may lead to new burdens or inefficiencies. Others do not seem to offer any reduction in costs or errors, but instead may permit a closer monitoring of cost or error components. The first column gives a short description of the CATI feature, the second two columns describe possible changes in costs or errors, and the next two comments on the ability to measure costs or errors. The column to the far right contains some notes on research that are needed to improve our methods under CATI. There are several different categories of features that are discussed in the tables. Table 1 reviews the facilities of questionnaires on CATI systems—how features of the nature of questionnaire setup can impact on the cost and error structure of surveys. Table 2 presents the features of the sample administration and sample control procedure. Table 3 examines supervisory and monitoring capabilities, and Table 4 examines coding.

The remainder of the article discusses the highlights of those tables, concentrating on the last column. Of chief interest is the ability to reduce costs or errors of surveys through the use of CATI technology. For many sources of error, such an outcome appears to be unlikely at this time. However, lacking a reduction of errors or costs, researchers would welcome the ability to measure them. For example, although CATI systems themselves may not reduce the effects that interviewers have on respondents, the ease of interpenetrating the sample across interviewers will permit routine estimates of interviewer variarance. First, it is useful to note that our expectation is that cost savings from CATI will arise primarily in the sample administration procedures and in the reduction of post-survey processing. That is, savings arise from transforming multiple activities into a smaller number of steps. On the other hand, opportunities for measurement of error arise through randomization facilities of CATI systems and through monitoring of survey processing; but by far, the clearest statements of change from CATI to non-CATI concern the measurement of errors. There are guesses on reduction of

TABLE 1
Changes in Measurement of Costs and Errors Related to Questionnaire Facilities

Feature	Changes In		Measurement Of		Needed Research and Development
	Costs	Error	Costs	Error	
Simple, Interactive Q'naire setup	computer file editing replaces cutting and pasting drafts but data cleaning specifications needed for setup	last minute skip changes can be added			software to test logic of q'naire program
Quick change of q'naire program	elimination of mid study memos	unnoticed errors can be quickly removed		count of changes to program may assist later cost estimation	interpretive vs. compiled design
Specification of valid codes	post-survey data processing reduced	immediate correction on interviewer error		counts of interviewer keying errors	how to announce error to interviewer
Ability to accept multiple entries per screen (e.g. grids)		reduce internal inconsistency of responses			best form of presentation of multiple items on screen
480 or more characters for open-ended response	immediate machine readable text	sensitivity to typing ability	timings on open-ended questions	text analysis of open-ended responses	machine coding of text responses
Alteration of question wording during interview ("How old is Harry?")	avoidance of multiple forms of the questionnaire	reduced interviewer errors in wording		response bias, variance experiments	balance of machine burden and interviewer error

Complex logic statements for skip logic		reduced missing data due to interviewer skip errors			likelihood of interviewer disorientation
Skip logic on text responses		reduced skip errors			level of machine load for text operations
Randomization of order of questions and/or responses		change of bias due to single form to variable error		context effects, order effects, response variance	routine randomization to measure response variance
Backward movement in the q'naire				number of respondent or interviewer errors	amount of interviewer freedom to move about q'naire
Movement through hierarchical data structure (e.g., persons in a family)	immediacy of movement may reduce interview time	reduce misplacement of data in subunits			nature of interface to interviewer, how to signal movement
Automatic review of text responses and comments	editing reduced to those questions in need of edit				
Hardcopy form of q'naire	reduced duplication costs				how to present complex q'naires in written form
Portability of question sequences in machine-readable form	reduction of marginal costs of repli cation		storage of timing estimates would permit better costing		alternative program designs to facilitate portability

TABLE 2
Changes in Measurement of Costs and Errors Related to Administration and Control Features

Feature	Changes In		Measurement Of		Needed Research and Development
	Costs	Error	Costs	Error	
Sample Generation On-Line	elimination of separate sampling function during survey	no lost coversheets	records of sample over days of the survey period	sequential sampling possible	desirable record keeping systems to study survey processing of sample numbers
Programmable callback scheduling rules	decreased number of calls per interview possible	effort at reaching each number can be specified by the researcher	collection of data on call patterns	maximal reduction of nonresponse conditional on number of callbacks specified	form of algorithms to guide callback schedule
Random assignment of sample numbers to interviewers				interviewer variance	handling of appointments refusal conversions
Reports on status of all sample numbers			average time spent on cases could be recorded	effort could be tailored to desired nonresponse rate	
Appointment lists for interviewers maintained on-line	no written record needed	machine reminders of appointment time possible	relative time spent on appointment cases		balance of machine work and human decisions
Supervisory ability to move sample cases to different statuses	machine analog of shuffling coversheets	machine tracing capabilities on cases			balance of machine activities and human activities
Automatic dialing of sample numbers	machine can pulse out ten digits per second	dialing errors virtually eliminated		can measure ratio of wrong connnections to misdialings	protocol for requesting a new number

TABLE 3

Changes in Measurement of Costs and Errors Related to Supervisory and Monitoring Capabilities

Feature	Changes In		Measurement Of		Needed Research and Development
	Costs	Error	Costs	Error	
Supervisor ability to observe any interviewer's screen			observation of how active interviewers are		
Machine based monitoring procedures		feedback from monitor can be numerical and personal		numerical monitoring data can be used in analysis	sampling procedures storage of data monitoring codes
On-line reliability coding of interview data				recording variability across interviewers	software to facilitate simultaneous recording of a case, selection of cases
Ability to review quickly completed cases	reduced search time for case examination	review of single interviewer's work easier			

TABLE 4
Changes in Measurement of Costs and Errors Related to Coding Capabilities

Feature	Changes In		Measurement Of		Needed Research and Development
	Costs	Error	Costs	Error	
Machine-readable text	reduced handling of paper, ease of 2nd coding efforts for later analysis	future machine coding capabilities	timings could be kept on coding	randomization to coders, automatic assignment to check coding	
Alteration of precoded categories based on initial responses	permits reduction of followup studies	refines measurement during survey		if done on replicates permits variance estimates	feasibility of within survey changes
Machine-readable comments on questions (by interviewer or respondent)	elminates need for coder review of all comments	possible to incorporate comments in data analysis			appropriate researcher use of comments

errors, although we do not even have any accounting systems yet to demonstrate the reduction of the errors.

One of the basic features in moving from non-CATI work to CATI is the ability to have machine readable text from open answers (item 5, Table 1). The fact that open-ended question responses are stored in this way permits new solutions to old problems. Two specific observations are pertinent. First—since in most systems each open response is a separate record—it is reasonable to sort the records by question and to permit coders to specialize on a small number of questions. That is, a single coder could code answers to a procedure that was not possible with paper questionnaires. This might permit coders to gain more experience with a single open-ended question and possibly reduce some coding errors. Conversely, it may be argued that such procedure limits the coder's knowledge of other responses given in the same interview and thus produces poorer coding. Second, the machine readable nature of text responses can alter the manner in which secondary analysts of survey data do their work. With archived machine text files, secondary coding of the data can be accomplished to tailor the coding to the particular needs of the data analyst. This form of data storage also permits new forms of summarization of text responses. In the near future, machine-driven text analysis systems could offer the researcher counts of word phrases, keywords, and separate ideas that now are not coded due to the costs of multidimensional coding of single responses.

The last item in Table 1 concerns the facility for backward movement in the questionnaire. Here, the issue is how much the researcher should allow the interviewer to move around the questionnaire. Early work on CATI by Rustemeyer counted the frequency of backing up during an interview. There are few sets of data like that, but they are badly needed now in the early stages of CATI development in order to measure the behaviors of interviewers during the questioning. This capabililty depends on a "log file" in the system, recording detailed key stroke level data.

An example of a general problem found in Table 2 is the first item. The change from a survey that has a single respondent per household to one that has more than one respondent per household is the simplest example of a hierarchical data structure, but one can think of many others where there are some types of records that are clustered, nested, or dependent in some sense. Such a data structure presents new problems to the interviewer during the interview. For example, the interviewer might learn in the middle of the interview from the second son in

a household that the disease mentioned for the head of household was wrong. Something else applies there, but the interviewer is deep into the data structure and has to get back to the head of the household and change his disease. What should we do? How do we tell the interviewer to do that? What is the most logical and least error-prone way of doing that? This is not a problem that will be solved by machine alone; indeed, the most difficult problem is the design of the interaction between interviewer and respondent to minimize response errors and missing data.

The third item in Table 2 results from our ability to construct a CATI questionnaire that is extremely complex. The Nicholls presentation of a flowchart form is also one that we use; indeed, such a device is more important as a questionnaire becomes more complex. The problem is, then, how does the researcher represent such a complex questionnaire in written form? From a different perspective, what is the appropriate review mechanism for OMB on disk-stored CATI questionnaires? We are actually using the flowchart as one level of documentation and using question units as another level of documentation; as questionnaires become more and more complex, it will become more and more difficult to put the entire questionnaire into the conventional paper format. It is likely that the appendices of survey monographs describing surveys are going to change radically over time.

The last item in Table 2 discusses one potential reward of on-line systems, and that is to make heavy investments in complicated but often used measurement forms and to transport those across organizations (such as measurement of occupation). Such investment might produce many sets of different questions—each set appropriate for a different subpopulation, but all attempting to measure the same characteristic. Such measurement procedures were unacceptable with paper questionnaires because of the complexity of interviewer and coder instructions that would be required for their implementation. Once an organization made such an investment, it would be most desirable to have the ability to transfer such measurement procedures across organizations. The coordination of CATI systems that would permit such transferability has not yet been achieved but is a goal that deserves our attention.

By far, the largest cost savings in CATI appear to be in the sample administration and control functions. There are, however, several unanswered questions in that area. It has been our view that the easiest

portion of an on-line system is the presentation of questions and movement between questions following the survey's skip logic. The more difficult procedures are choosing the next number to be dialed, scheduling callbacks, and doing all of the sample administration work. This part of the system is complex partially because we have not clearly specified how it is done now. Supervisors somehow make decisions that are not documented on who is going to call what when. We can exert some influence over that, but we do not know how often there are deviations from those guidelines. It is difficult to design optimal algorithm because we do not have the data to judge optimal calling, especially that conditional on a previous calling sequence. So we have to develop data bases and gradually iterate into some desirable algorithm. There needs to be work on the general application of sample administration.

We have spent a lot of time attempting to take advantage of the randomization capabilities of on-line systems (item 3, Table 2). We have a lot of interest in measuring interviewer variability, to assign interviewers, in essence, random subsamples of an overall sample. There are, however, unanswered problems in that area as well, mainly statistical problems about handling appointments and refusal conversions cases that are typically assigned to particular interviewers and, if possible, to refusal converter interviewers that are skilled at persuading people to cooperate. These departures from randomized assignment create problems in the estimation of interviewer variability.

Monitoring can be used for several purposes (item 2-3, Table 3). It can be used to give feedback to interviewers, regarding their performance, and thus to reduce errors on the same survey. It can also be used to collect empirical data on the nature of the errors without any feedback. One can imagine a system machine driven that would sample interviews as they were ongoing for monitoring attention. It could search for cases that were approaching a certain question sequence that the researcher wanted heavily monitored. One could create, from such a system—in addition to a vector of data items—a companion vector of characteristics of the interviewer's delivery of the question on that particular question, the nature of their feedback to the respondent, and so on. Thus, on a subsample of the cases, an auxiliary set of variables could be used in analysis and used in the evaluation of interviewers. By "on-line reliability coding," we mean having a monitor sit at another

terminal following the path of the interview that is going on at the moment and put down what he or she thinks the respondent is saying at each point. The goal is to do this on a subsample of cases as well, and to provide new estimates of recording reliability. Both of those are examples of ways to begin to collect numblers on the nonsampling error components of telephone surveys that will allow us to point out the change in the error structure on CATI systems, so that we would be looking at both changes in the cost and changes in the error structure.

SUMMARY AND CONCLUSIONS

There are four observations of relevance in a summary of this discussion. The first is that it seems that most of the changes in moving from non-CATI work to CATI work will arise not in the reduction of costs but in changes and potential changes in the error structure; and that comment itself has to be conditional on the fact that the cost structure is changing all the time, depending on the nature of the hardware purchased, so it will be continuously in flux; but the immediate gains likely will not be in terms of cost, but rather in the nature of the error structure.

Second, to evaluate effectively this change from telephone surveys without computer assistance to CATI work, we have to start gathering empirical data on the magnitude of of errors. We can easily collect datas on the magnitude of costs, but we have not been collecting it on the magnitude of errors, and we have to take advantage of the randomization capabilities—the monitoring capabilities of CATI systems—to start collecting those data.

Third, a point mentioned before is that we are just beginning to use this new technology and we are doing it to solve all our old problems. There are others we can attempt to solve with this new technology. It has been suggested that we use systems to control the pace of the interviewer. We can start experimenting with new question forms that we never look at before. There are also new problems created by the use of the technology itself. We should be concerned about interviewer fatigue from working at the screen for long hours and problems of interviewer confusion about the current location in a complex questionnaire.

The fourth point is that the recurring theme throughout all of these discussions is the balance between stripping activities that the interviewer did previously away from the interviewer and letting the machine do them. How far do we go in skip logic and consistency checking before

we start to burden the machine or make the interviewer's task an uninteresting one? It is a problem that will be dependent on each system's solution to CATI, but it will always be an issue that arises in our decisions.

Robert M. Groves is a Senior Study Director of the University of Michigan Survey Research Center and an Associate Professor in the department of Sociology. In addition to his interest in CATI, he is involved in research on dual frame, mixed mode designs, that combine face-to-face and telephone interview methods, and the effects of survey procedures on various nonsampling errors.