DISCUSSION

David Judkins, Westat
1650 Research Blvd., Rockville, MD 20850

KEY WORDS: Nonresponse Bias, Quantiles, Small-area Estimation

I want to thank the authors for a very interesting set of papers. As I reviewed their papers before the conference I was struck by the diversity of the topics. At first, they seemed to have very little in common. The first is a demographic paper on standards for growth in children that will be used in doctors' offices around the world. The second is an economic paper on a new time series of income tax data. The users will probably be confined to economists at government offices, think tanks, and universities. The third is a survey methods paper reporting on possible nonresponse bias in a specific survey. Two groups will be interested in it: survey methodologists like myself and economists who use the survey. The fourth paper is on uses of auxiliary data in improving estimates from a survey of businesses in Canada. It shares the same audience of survey methodologists and economists using the survey. The fifth paper is on methods for quantile estimation for small areas from complex surveys. The material is rather technical, near the cutting edge of theoretical research for sample surveys and, hence, of immediate interest mostly to theoretically oriented government and academic applied statisticians.

Despite this rich variety of topics, levels of presentation, and intended audiences, there is a common thread that runs through the papers. They all have something to do with percentiles or quantiles of probability distributions. In the paper presented by Roche, we heard how new and improved **normative** quantiles have been prepared on growth of children. In the paper presented by Scheuren, we heard how estimated quantiles from different cross-sectional reports are being standardized so that they may be compared across years. In the paper presented by Woodburn, we heard how quantile-quantile (Q-Q) charts were used to help decide whether volunteers were substantially different from non-volunteers. In the paper presented by Lee, we heard how the "regression quantile method" was used to identify outliers in auxiliary data to improve its utility in estimation from a sample survey of businesses. In the paper presented by Kovacevic, we heard an in-depth discussion of how to estimate quantiles for small areas using both design-based and model-based techniques.

A common thread, but a rather thin one to put much weight on. Rather than emphasize this commonality any more, I wish to comment on several aspects of the papers that are nearest to my area of specialty. In particular, I was very interested to read the paper by Woodburn and Heeringa.

Research into nonresponse bias is a very important topic that we all agree should be done more often. The reason that it isn't done more often is that this type of research is very hard. It is, naturally enough, rather difficult to analyze people who won't allow themselves to be observed. This type of project is most successful when the initial response rate is very bad (low) or when the sample can be matched to a list with rich auxiliary data. If the initial response rate is bad, then experiments can be devised to boost the response rate on a subsample which then allows analysis. This particular project was successful because of matching. Of course, exact matching is usually a very difficult problem as well. (I might point out here that one of the other presenters, Dr. Scheuren, organized an excellent conference on exact matching a few years ago, the proceedings of which I presume are still available.) Besides problems of keying errors, there is usually substantial red tape to cut through. The authors succeeded because of the cooperation of the Federal Reserve Board and the IRS.

An early version of this paper concerned me with it's lack of attention to power considerations. The sample sizes were too small for traditional hypothesis tests to detect any but the strongest of effects. It serves no useful purpose to conduct an investigation with small sample sizes and then conclude that there are no significant differences. It is even worse to then conclude from this lack of a negative finding that the data are alright. Any data set with less than perfect response should be presumed to be subject to potentially serious nonresponse bias. Certainly, a survey with a response rate under 10 percent must be presumed guilty until proven innocent. The only way to demonstrate that nonresponse is not a serious problem is to have a sample size that provides confidence intervals (95 percent or better) that are tight enough that no analytically significant differences lie outside them.

In the presentation we heard today, however, the t-tests were replaced by a powerful graphical technique that made clear the differences between volunteers and nonvolunteers. The authors are to be commended for this insightful analysis.

In the paper by Kovacevic, I would like to draw your attention to his nearly neutral treatment of design-based and model-based inference methods. I thought that it was a balanced approach that contributes to bridging the chasm between those who believe that designed randomization provides the basis for statistical inference and those who believe that models provide that basis.

Near the end of his paper, he mentions how difficult it is to estimate variances on small area estimates of cdf's and quantiles. He rightly points out that resampling methods are problematic because they usually involve discarding some of the data. When data are discarded, the small-area statistic can easily become undefined due to a zero denominator. I would draw his attention, however, to a resampling method due to Robert Fay that doesn't involve discarding any data. Instead, weights are gently perturbed in a modified balanced repeated half-sample scheme.[1]

The work done by Lee and Croal is a good example of how to proceed when considering the use of auxiliary data in estimation. Researchers on other surveys would do well to repeat much the same sorts of analysis with their own database. Naturally, however, one must be careful about applying their results to other surveys. The results themselves may be very dependent on the variables and cases examined.

In conclusion, I wish to thank the authors again for their excellent presentations and to thank the chair for the opportunity to make these few brief comments.

---

[1] David Judkins, "Modified balanced repeated replications," *Proceedings of the American Statistical Association Section on Survey Research Methods* (1987), p. 492-495.

METHODOLOGICAL ISSUES IN CAPI

Mick Couper, Robert M. Groves, The University of Michigan,
Carol Kosary, Arbitron Ratings Company
Mick Couper, 2104 ISR, P.O. Box 1248, Ann Arbor, MI 48106

KEY WORDS: CAPI, Computer assisted personal interviewing, data collection

Introduction

Computer assisted personal interviewing (CAPI) is a survey design option in its infancy. As it is currently conceptualized, it is a merging of two technologies: the face to face interview and computer assisted telephone interviewing (CATI). This marriage of two methods tends, in our opinion, to be viewed as a welcomed, albeit arranged union. This paper focuses on some of the properties of this union which may require more careful thought and demand a research and development program. The paper begins with a tentative definition of CAPI. In addition to looking briefly at certain features of the new technology that require research, we suggest some approaches to such research, and discuss potential problems in conducting research on CAPI.

Toward a definition of CAPI

Despite the lack of an explicit statement of what CAPI entails, the acronym has already become common parlance among survey practitioners. It appears that a number of organizations involved with CAPI use the term to refer to the interviewing phase of the survey process only. It is felt that this current view of CAPI is too limited, and should be expanded to include computer assistance in all phases of the data collection process. Following the approach of Nicholls and Groves (1985) on CATI, in Exhibit 1 some of the features that should be considered part of CAPI are outlined. In its most basic form, CAPI uses portable computers to present survey questions for interviewer administration, receive and store responses, and assemble data to send to a central office.

Most work to date has focused on the interviewing functions of CAPI. However, as progress is made towards a paperless data collection environment, the other features of CAPI outlined in Exhibit 1 will receive increasing attention. CAPI should not be seen simply as CATI on a laptop. It appears as if the greatest advantages of CAPI lie in the front and back ends of the process rather than the interviewing phase.

Given the current state of CAPI research and development, a number of areas of the new technology are identified that are ripe for methodological research. Several of these components are discussed here: hardware, software, supervision and control, interviewers, respondents, interviewer-respondent interaction, and the interview environment. Some of the issues faced by those investigating the feasibility of CAPI are also briefly outlined.

Hardware

Much of the focus regarding CAPI to date has been on the choice of appropriate hardware. This is probably the most crucial component in the success or failure of CAPI, yet at the same time it is the most volatile.

It appears that there are two general approaches to the question of hardware, each serving different needs. Those organizations conducting lengthy interviews in the respondent's home are focusing on laptop computers (see NCHS, 1988; Sebestik, et al., 1988; Rothschild and Wilson, 1988). Others whose needs include in-store pricings and short interviews which are often conducted on the respondent's doorstep are looking towards handheld computers (see for example Bemelmans-Spork and Sikkel, 1985a, 1985b; Netherlands CBS, 1987). These approaches raise very different issues for research. The use of CAPI hardware for multiple purposes further complicates the choice and evaluation of hardware. Some common concerns that have been raised by interviewers are issues of weight and/or bulk, screen readability, speed of operation, and battery power.

Despite the range of research issues on hardware, to our knowledge, no between-machine or between-component analyses of hardware have been undertaken. Given the volatility of computer technology, we feel it is more useful to focus research efforts on particular components of the hardware (screen, keyboard, weight, battery life, speed, etc.) rather than on whole machines. In this way a set of optimal hardware features can be identified to meet specific study or interviewer requirements.

Software

The scope of software issues relates to the conceptualization of CAPI. Currently the most common approach involves the transfer of CATI software and functions to CAPI. This should be viewed only as an interim step in the development of CAPI. Generally such

software developments have involved only one aspect of CAPI, namely the core interviewing functions. Other software functions that are unique to CAPI (such as telecommunication, transmission of cases, sample administration, etc.) need to receive more attention. The development of software to perform these functions will be a key factor in the successful implementation of CAPI and its acceptance by field staff.

Supervision and control

The electronic and decentralized nature of CAPI creates the potential for multiple copies of cases (complete or otherwise) to be in circulation. Decisions need to be made regarding the direction, frequency and content of the information flow between the interviewer, supervisor and central office. Protocols for the management and control of cases need to be developed. One issue is whether (or to what extent) interviewers and supervisors should be allowed to access and edit cases after completion. Backup systems are needed for a variety of contingencies such as machine failures at the interviewer or supervisor level, or problems with the transmission of sample cases and/or data.

The role of the supervisor (particularly at the regional level) also needs to be redefined for CAPI. The functions of the supervisor may change as a result of direct transmissions to and from a central location and the potential for electronic monitoring of data quality and interviewer productivity from the central office. The role of the supervisor before, during and after data collection needs to be re-evaluated. CAPI permits supervisory oversight not permitted before in field interviewing, but also takes other functions away from interviewers. The successful implementation of CAPI requires a degree of organizational change. The extent and form that this restructuring may take is an area that requires research.

Interviewers

It is expected that the greatest impact of CAPI will be felt by the interviewers using the computers. In most cases the introduction of CAPI will involve the imposition of a new technology on an existing labor force. This new approach to data collection is expected to substantially alter the nature of the work currently done by field interviewers. This has both psychological and physical implications for the interviewer.

Some of the issues that should be addressed at the level of the interviewer include (a) computer attitudes or computerphobia, (b) physical capacity for work with laptop and/or handheld computers, (c) training considerations, (d) computer literacy and related issues of competence, (e) use of machines for purposes other than data collection, and (f) concerns of victimization, crime and damage. Interviewers' reactions and acceptance of CAPI is expected to be a major determinant of the success of the new technique.

Respondents

The impact of CAPI on respondents has yet to be fully explored. Respondents' expectations may be affected by the introduction and use of a computer. Both positive and negative impacts may be possible. Some potential benefits of CAPI may be an increased perception of confidentiality and increased interest due to the novelty of the new technology. Potential drawbacks include a perceived loss of privacy and fear of computers. Initial findings on respondent reactions are encouraging, and it appears that the impact of the computer on the respondent is slight (see for example Groves and Nicholls, 1986; Bemelmans-Spork and Sikkel, 1985). However, in most studies of respondent reactions, data were collected by the interviewer during or immediately after administration of the CAPI instrument, raising questions of socially desirable responses. The potential effects of CAPI on response rates, reluctance to respond, response bias and issues of confidentiality all require further attention.

Respondent-interviewer interaction

Unlike CATI, in which the computer is invisible to the respondent, the presence of the computer for CAPI is expected to have an impact on the nature of the interaction. However, what this effect may be is unknown at this stage. The effect of CAPI on eye contact and seating arrangements are just two of the issues that need to be examined. A consistent finding in early CAPI studies (see Birkett, 1988; Waterton and Duffy, 1984) is that interviews take longer using the computer than with paper and pencil. Such time delays using CAPI are usually attributed to problems with the hardware and or software being used. Whether this will change over time as interviewers become more familiar with the system or whether it is an inherent feature of CAPI is not yet clear. Evidence from CATI is also that fewer interviewer comments are received with computer assisted methods, presumably because of the extra keystrokes required (see Groves and Nicholls, 1986). This may be further reduced in CAPI because of the possibility of the respondent seeing what the interviewer is entering on the screen.

Environment

Many of the impacts of the external environment on computer assisted data collection are defined away under CATI. Interviewers are often located in customized facilities that optimize conditions for computer use. In contrast, with CAPI one has little control over the environment in which the computer is to be used. Factors such as temperature, lighting, and humidity all have a potential impact on the effectiveness of CAPI. In those surveys in which interviews are conducted outside the respondent's home, the potential impact of the environment on CAPI becomes even more critical, not only in terms of hardware failure, but also in terms of the interviewers' ability to operate a computer under adverse climatic conditions. The impact of these factors on the implementation of CAPI has yet to be explored. In addition, with CAPI, all the distractions of face to face interviewing (children, TV, pets, etc.) will exist. These may all serve to magnify the problem of segmentation currently experienced in CATI. Segmentation refers to the tendency for interviewers to get disoriented in complex computerized instruments because of the lack of a mental picture of the instrument's overall structure (see Groves and Nicholls, 1985).

Research issues

These are just some of the numerous issues facing those working with CAPI. To date what little research has been conducted on CAPI has been limited to feasibility studies involving qualitative evaluations of CAPI. Generally these have been small scale field tests with a single hardware/software configuration rather than experimental designs of alternative approaches. Although a few of studies have compared CAPI with paper-and-pencil data collection, the results have thus far been inconclusive. Birkett (1988) found a higher error rate for CAPI, and attributed it to poor training and hardware/software problems. A similar conclusion was reached by NCHS. As yet there is no empirical evidence to show that CAPI is better (in terms of costs and/or data quality) than traditional paper-and-pencil methods. There are a number of problems facing those planning research on CAPI, which may in part explain the lack of research to date.

Probably the greatest limiting factor in CAPI research is one of cost. CAPI is an expensive undertaking, involving substantial investments in equipment, software development, training and staff time. Few organizations can afford to mount large-scale tests of CAPI in a pre-production mode. For this reason, no studies have been conducted comparing alternative hardware or software configurations for CAPI, and only a few have compared CAPI data collection to traditional methods.

Another question concerning research on CAPI relates to the basis of comparisons that should be used for evaluating CAPI. Arguments can be made for comparing CAPI to both paper-and-pencil and CATI data collection. For example, CAPI can be expected to produce errors that differ from those often found in paper-and-pencil surveys. Item missing data is generally restricted in CAPI, whereas keying errors may be more prevalent (see Sebestik, et al., 1988).

A further factor complicating CAPI research lies in the fact that interviewer reactions to and skills with the new technology may change over time. The novelty of the new method for both interviewer and respondent may wear off, resulting in an increasingly negative perception of CAPI. On the other hand, speed and quality may well improve over time as interviewers become more familiar with using the system. Only a longitudinal study can address these issues.

Related to this issue is the possibility that the generally positive reactions to CAPI on the part of interviewers and respondents may in part be due to a Hawthorne effect. The fact that they are the focus of attention, and have been selected to evaluate a new system may lead interviewers to more favorable reviews than may generally be expected. Similarly, the presence of observers to evaluate respondent reactions and the interviewer-respondent interaction in CAPI may well affect the process itself. Many of the measures used to evaluate CAPI may thus be affected by the interventions needed to obtain such measures.

Despite the limited empirical evidence in favor of CAPI, many organizations are adopting the new technology on the assumption that the problems currently experienced will be overcome. However, nobody has yet undertaken a comprehensive empirical analysis of the advantages and disadvantages of CAPI relative to other techniques of data collection. Initial expectations of CAPI are that the major advantages of the new approach may lie in data quality and turnaround time, while the biggest disadvantage is one of cost. However, these issues have not yet been adequately addressed empirically.

Research Approaches

Given the limited research that has been done on CAPI to date, and the problems that inhibit such research, there are a number of ways in which some of these research issues can be addressed. Because of the complexity of CAPI, and the number of issues that impact on its effectiveness, a single

comprehensive test to address multiple issues has limited feasibility. Given the volatility of CAPI hardware, it is felt that tests using a single machine cannot address the many issues related to various components of the hardware. Instead it is proposed that each of the components of CAPI discussed in this paper can be addressed separately. Such more modest tests may well provide the data necessary to resolve particular issues within the general domain of CAPI research. The use of multiple machine comparisons (where feasible) is also strongly advocated, as is the use of experimental designs to compare various alternatives.

A number of research approaches can be adopted to address these issues. These include laboratory tests, field tests, surveys of interviewers and respondents, and simulations.

Laboratory tests offer one way of countering the exorbitant costs of mounting full-scale field tests of CAPI. Many of the issues relating to hardware, software and interviewer reactions to CAPI can be addressed in the laboratory. This approach also permits the evaluation of alternative hardware and software configurations with fairly limited resources. The drawback of laboratory studies lies in the generally small sample sizes used, and the short duration of the tests. Referring to wrgonomic testing in general, Litekwa (1987) notes that ten minute trials are not valid tests. Despite these limitations, laboratory tests are a useful tool for choosing between various options prior to a larger field test.

Where field studies can be launched, the use of experimental designs rather than feasibility studies is strongly advocated. To date, no field test has compared alternative CAPI designs or configurations. The development and testing of specific hypotheses relating to various components of CAPI in an experimental setting will advance our understanding of the new approach.

Another approach to gathering data on the potential impact of CAPI is through surveys of both interviewers and the general public. To our knowledge, no survey has yet addressed the question of attitudes toward computers using a national probability sample. It is felt that respondents' reactions to being interviewed with CAPI should be measured independently of the CAPI interview (preferably by a third party), to reduce the possibility of social desirability effects. Interviewers' expectations regarding the new technology can also be explored both before and after exposure to CAPI. Field interviewers in many organizations tend to be older females, the very group that is expected to be most anxious about using computers (see Gilroy and Desai, 1986). It is thus important that their attitudes to CAPI be assessed.

Findings from research in other fields can also contribute to our understanding of the process of CAPI data collection. Generally, this research has not addressed the issue of computerized data collection or the use of laptop of handheld computers directly, but many findings have indirect relevance to CAPI. There is a substantial body of research in the field of ergonomics that deals with the question of manual materials handling and physical capacity to lift, hold and carry various objects (see for example Haisman, 1988). Similarly, kinesiology deals with the physical capacity to handle various tasks. Questions of screen readability, keyboard configuration, presentation of information on computers screens, speed of operation, reaction time, and other aspects of human-computer interaction have all been addressed in both the ergonomics and the computer science literature (see for example Grandjean, 1988) . The field of environmental architecture has looked at the effect of lighting and other environmental conditions on the operation of computers. Finally, research on CATI and on paper-and-pencil interviewing has relevance to the CAPI. As a hybrid method, CAPI has features in common with both these approaches, and we can learn from research conducted using other data collection approaches.

Conclusion

This paper has been written in the hope that it will stimulate those who are currently involved in CAPI to undertake the basic research needed to evaluate the efficacy of this new method. There is a large set of research yet to be undertaken on this new technology. However, for those with pressing operational needs, timely research is needed on the effect of CAPI on costs and errors in survey research. This paper is intended to make this latter group aware of the complexity of this new technique and alert them to some of the potential dangers of using a method that has yet to prove itself. CAPI holds a great deal of promise, but we should approach this new technique with caution. It has yet to be demonstrated to be superior to current paper-and-pencil approaches to data collection. We need to take a critical look at this method before hailing it as the panacea for all our data collection problems.

Exhibit 1

Features of CAPI

Hardware/software to permit telecommunication - modems, telephone access, encrypting or packing software, telecommunication software

Telecommunication for remote problem diagnosis - ability for central office to obtain software control of interviewer's machine within telecommunicaton session

Transfer of sample cases to and from interviewer - sending of work assignments from central office or supervisor to interviewers; sending of completed assignments from interviewer to central office or supervisor

Telecommunication of status, case data - ability of host machine to query interviewers machines to monitor progress of survey activities

Messaging systems - ability for interviewers to send communications on field problems or progress to supervisors or central office

Sample administration - ability to list and sort assigned cases, schedule work on assignments, and monitor progress on assignments using software assistance

Transfer of cases to different interviewers - ability to reassign cases from one interviewer to another, protocols to avoid duplicate assignment or lost cases

CATI-like data entry functions - presentation of questions, edit checks, skips, comments, online help functions

Preparation of data sets - merging of cases into data set for analysis; editing and/or coding; translation into output data set

References

Bemelmans-Spork, M.E.J. and Sikkel, D. 1985. "Observations of prices with hand-held computers." Statistical Journal of the United Nations Economic Commission for Europe, 3: 153-160.

Bemelmans-Spork, M.E.J., Keller, W. and Sikkel, D. 1985. "Data collection with hand-held computers." Proceedings of the 45th session, International Statistics Institute, Book III, topic 18.3.

Bemelmans-Spork, M.E.J., Keller, W. and Sikkel, D. 1985. "Use of handheld microcomputers in the Netherlands Central Bureau of Statistics." Proceedings of the Bureau of the Census First Annual Research Conference. Washington: Bureau of the Census, pp. 47-52.

Birkett, N.J. 1988. "Computer-aided personal interviewing: A new technique for data collection in epidemiologic surveys." American Journal of Epidemiology, 127 (3): 684-690.

Drury, C.G. 1987. "Hand-held computers for ergonomics data collection." Applied Ergonomics, 18 (2): 90-94.

Gilroy, F.D. and Desai, H.B. 1986. "Computer anxiety: sex, race and age." International Journal of Man-Machine Studies, 25: 711-719.

Grandjean, E. 1988. Fitting the Task to the Man. (4th ed.). London: Taylor and Francis.

Groves, R.M. and Nicholls II, W.L. 1986. "The status of computer-assisted telephone interviewing: Part II - Data quality issues." Journal of Official Statistics, 2 (2): 117-134.

Haisman, M.F. 1988. "Determinants of load carrying ability." Applied Ergonomics, 19 (2): 111-121.

Litekwa, J. 1987. "Questionable assumptions and misconceptions in ergonomic research on human-computer interaction: A discussion." Pp. 337-343 in Salvendy, G., Sauter, S.L., and Hurrell, J.J. (eds.). Social, Ergonomic and Stress Aspects of Work with Computers. Amsterdam: Elsevier.

Lyberg, L. 1985. "Plans for computer assisted data collection at Statistics Sweden." Proceedings of the 45th Session, International Statistics Institute, Book III, Topic 18.2.

National Center for Health Statistics and Bureau of the Census. 1988. Report of the 1987 Automated National Health Interview Survey Feasibility Study. Hyattsville: National Center for Health Statistics (Working Paper Series no. 32).

Netherlands Central Bureau for Statistics. 1987. Automation in Survey Processing. The Hague: Government Printer.

Nicholls II, W.L. 1989. The Impact of High Technology on Data Collection. Washington: Bureau of the Census (CATI Research Report No. GEN-1).

Nicholls II, W.L. and Groves, R.M. 1986. "The status of computer-assisted telephone interviewing: Part I - Introduction and impact on cost and timeliness of survey data." Journal of Official Statistics, 2 (2): 93-115.

Ochoa, L.H., Ortuzar, J., and Willard, J.L. 1988. "Procesamiento interactivo de encuestas complejas con microcomputadores portatiles: un experimento en Guatemala." International Union for the Scientific Study of Population, Seminar on Collection and Processing of Demographic Data in Latin America.

Popper, R. et al. 1988. "Use of pocket computers for self-administration of cognitive tests in the field." Behavior Research Methods, Instruments, and Computers, 20 (5): 481-484.

Rothschild, B.B. and Wilson, L.B. 1988. "Nationwide Food Consumption Survey 1987: A landmark personal interview survey using laptop computers." Proceedings of the Bureau of the Census Fourth Annual Research Conference. Washington: Bureau of the Census, pp. 47-52.

Sebastik, J., Zelon, H., DeWitt, D., O'Reilly, J.M., and McGowan, K. 1988. "Initial experiences with CAPI." Proceedings of the Bureau of the Census Fourth Annual Research Conference. Washington: Bureau of the Census, pp. 47-52.

Statistics Sweden. 1989. Computer-assisted Data Collection in the Labour Force Surveys; A Report of Some Technical Tests. Örebro: Statistics Sweden.

van Bastelaer, A., Kerssemakers, F., and Sikkel, D. 1988. "Data collection with hand-held computers: contributions to questionnaire design." Journal of Official Statistics, 4 (2): 141-154.

Waterton, J.J. and Duffy, J.C. 1984. "A comparison of computer interviewing techniques and traditional methods for the collection of self-report alcohol consumption data in a field survey." International Statistical Review, 52 (2): 173-182.