

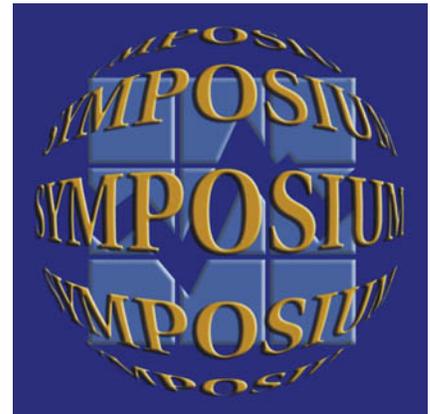


Catalogue no. 11-522-XIE

**Statistics Canada International Symposium
Series - Proceedings**

**Symposium 2004: Innovative
Methods for Surveying
Difficult-to-reach Populations**

2004



THOUGHTS ON THE FUTURE OF SURVEYS

Norman M. Bradburn¹

1. INTRODUCTION

Almost thirty-five years ago I wrote a paper about what survey research might be like if there were developed something that was called the "information utility" (Bradburn, 1970). The then revolutionary idea was that the extension of coaxial cable to most households in the U.S. opened the possibility to send messages out of the household by means other than the telephone. On the analogy of electricity and telephones as public utilities, the ability to send information in textual and visual in addition to auditory mode from the household to external receivers, as well as the reverse, was conceptualized as an "information utility." The term did not catch on, partly, I think, because the vision was not turned into reality until many years afterward, and then not in the form that was envisioned.

Rereading the paper today, I find that I failed almost entirely to foresee what actually developed, but I think I did a reasonable job in pointing out the challenges that would have to be overcome if what was imagined by the information technology visionaries was to have a significant impact on the conduct of sample surveys. I don't feel particularly bad about not foreseeing what actually happened. Subsequent technological and market developments and other economic, social and political forces shape how technological possibilities that can be imagined by smart and creative people become realities. There are too many unknowns in this developmental process for anyone to be a very accurate forecaster.

At the recent ASA meetings in Toronto, Dean Judson of the U.S. Census Bureau described the last half of the 20th Century as the golden age of the sample survey. By this he meant that this was a period of technical improvement in surveys and a flowering of the use of surveys for social informational, program evaluation and marketing purposes. It was a period when surveys could achieve reasonably high response rates at acceptable, although growing, costs. He argued that the growing respondent resistance to surveys will mean that social information in the 21st Century will have to adopt new strategies, particularly blending different modes and types of data, such as administrative records, to come up with the best estimates of whatever we are trying to measure. New technologies may change our conception of how to do surveys.

Today I want to make another attempt to look at the future of survey research in the light of the incredible developments in information technology that we are in the midst of, developments that might usher in a new golden age. I am going to turn the question around, however, and start with the perennial requirements for conducting surveys and speculate on what new technologies might contribute to doing surveys differently. I want to emphasize that these are speculations because, while most of the technologies I will talk about exist in some form or another, some may not be any further along than the vision of an information utility was in 1970. This is a talk about what could happen, not what will happen.

I will organize the discussion about the steps in conducting a survey: sampling, selecting the mode for the survey, contacting the respondents, developing the questionnaire, and processing the data. Finally, I will comment on a few issues that do not fit neatly into this paradigm.

¹ NORC/ University of Chicago

2. SAMPLING

I start with the problem of selecting the sample. In order to limit the discussion, I will confine my attention to surveys where the unit of analysis is either households or individuals. I assume that no one in this audience is interested in what I call SLOPs (Self-selected Listener Opinion Polls), that is surveys in which the sample is self-selected. The sampling task is to define the universe from which the sample is to be selected and devise a probabilistic method for selecting a sample from the universe. Leaving aside the relatively simple case of list samples, the universes of interest are usually households within some geographic area, e.g. the continental United States, the city of Chicago, or Arlington County, Virginia and/or individuals living in those households. Currently the two methods for constructing the universe are 1) creating your own list by field listing and counting, that is sending out enumerators to examine the areas of interest and determine which addresses are actual households, or 2) using lists already constructed, e.g., commercially available address and telephone lists or logical lists, such as all possible 10-digit numbers. You can then either sample from the list directly or use random number generation to sample from a logical list. For universes in which we do not know the actual individuals, we are essentially limited by current technology to two types of lists – addresses and telephone numbers – as the basic sampling unit. From those starting points we then use questions of informants to eliminate non-members of the universe and locate the desired sample person or household.

The technology that is currently exciting interest as a new way of conducting surveys is the Internet. Aside from the temporary problem of household coverage, the principal barrier to using the Internet for surveys other than SLOPs is the sampling problem. There is no list of households or individuals with access to the Internet nor are there any developments on the horizon that will produce such a list. In 1970 I wrote that: "In order to realize the full advantage of the hypothesized technology, we shall inevitably move toward a registry of persons, or at least a registry of receiving units for the information to be sent via the information utility or both." (Bradburn, 1970).

As a country, the United States is still resisting the registry of persons, although we are perhaps inching toward it with fear and trembling. We do not have a list of devices capable of accessing the Internet, although I think the technology for constructing one exists. If all devices accessing the Internet had fixed unique IP addresses, it should, in principle, be possible to construct a list, or sample from a logical list. Until something analogous to telephone numbers exists for Internet accessible devices, we will not be able to do probability sampling using the Internet.

There are some interim solutions that exist or could come about in the relatively near future. The one that exists now is the method employed by Knowledge Networks, which, as far as I know, is the only organization that can do probability based surveys via the Internet. That solution starts with a RDD telephone sample, recruits sample households to participate in a panel, and places Web-TV boxes in the sample homes to give Internet access using a uniform platform. This method creates a probability sample of households with known characteristics of the universe. If one were worried about the bias against non-telephone households, the method could be extended by face-to-face recruiting of households without telephones. It is an expensive method, but it avoids the self-selection bias that plagues all the other Internet surveys

A technological development that would overcome some of the problems would be the combination of wireless telephony with Internet access and the unique assignment of telephone numbers to individuals independent of their geographic location. While the combination of wireless telephony and Internet access solves part of the problem in that it establishes the possibility of a logical list that could be sampled, the change to lifetime telephone numbers associated with unique individuals would truly revolutionize survey practice in that it would make the direct sampling individuals possible. On the other hand, the detachment of telephone or other ID numbers from geographic location would make geographic-based sampling difficult. The growth of Voice over Internet Protocols (VOIP) may make this a reality in the not to distant future.

Further in the future is Internet telephone, that is, direct video or auditory communication between computers. But if Internet telephony became common, it would end surveys as we know them unless the device sampling challenge can be met.

Another technology that may take on increasing importance as a result of homeland security issues is Geographic Information Systems (GIS). GIS technology is making great progress, and it already has many uses both for mapping addresses and for analytic purposes when survey data are linked to geographical data. If any of the

various individual registry options were combined with geographical tracking technology, the possibilities for sampling would be revolutionized. At the same time, some of the technological possibilities raise fundamental questions about privacy and civil liberties that would have to be solved before such systems could be employed.

To summarize, for the Internet or any other advanced IT based mode that may come along to make a major impact on survey research, there needs to be a method for constructing lists of the households or individuals that have access to the Internet and a method for contacting them directly through the Internet or by some other means. An example would be a list of all valid email addresses, and the ability to unduplicate them. If there were a common format for email addresses, such as exists for telephone numbers, one could create a logical list and do some form of RDD sampling of email addresses. How useful such a list would be, of course, would depend on the amount of coverage, but at least, one would be on the way in principle to being able to do probability sample with other technologies than the telephone, regular mail or personal visits.

3. SELECTING THE MODE FOR CONDUCTING THE SURVEY

The current modes for conducting surveys are face-to-face interviewing, telephone, paper-and-pencil self-administration and, increasingly, email and the web. The characteristics and relative advantages of face-to-face interviewing, telephone and paper-and-pencil modes have been the subject a vast amount of methodological research over the years (Sudman and Bradburn, 1974; Groves, et al. , 1988), but research on the modal characteristics of web-based surveys is in its infancy (See Dillman, 2000 for a summary of current knowledge). We no doubt will learn much more in the next decade.

The fascinating possibility for web-based surveys is the ability to use multi-media presentations of background material, innovative ways of asking questions, and new formats for recording answers, freeing researchers from the limitations of essentially text based methods of inquiry. While much more can be done with the existing state of technology, even more exciting new ways of conducting surveys could be developed if broadband communications were commonly available. One can imagine two-way video communication that would allow interviewers and respondents to establish, through electronic means, the type of communication that is now only possible with face-to-face interviewing. Future technological developments that would create even more realistic representations of the communicating parties might open up even greater possibilities for establishing the kind of rapport with respondents that that has been one of the strengths of face-to-face interviewing.

At the 1995 methodological conference in Bristol (Blyth, 1997) the talk was all about voice recognition as the technological development most likely to revolutionize the way surveys were done. Although synthesized speech on the part of the interviewer and voice recognition for simple response categories are becoming more widely used in corporate applications and in simple surveys, their full potential still seems to me to be far off. Progress in this domain, I believe, is hampered less by the technology of speech synthesis than by the limitations of our fundamental knowledge of language and cognition. Real progress will only be made when we know more about how meaning is encoded in language and how people comprehend questions and formulate answers.

In the very far future one can conceive of the development of what might be called interviewer bots, that is, electronic bots capable of simulating interviewers with enough flexibility and "intelligence" to respond meaningfully to respondents who comment or ask questions about the survey questions. Information technology research is very active in this area. I expect that if I were giving this talk again 30 years from now, it would be something that might be nearing a reality.

To sum up, the advent of web-based surveys opens up some interesting possibilities to alter drastically the way we present questionnaires by freeing us from the limitations of using only auditory or text-based modes. While visual materials have been used in some face-to-face surveys, they are difficult to prepare and difficult for interviewers to use. The integration of visual stimuli, particularly video materials, into questionnaires is a challenge to our habitual way of doing things. I hope we can rise to that challenge.

In the longer run, the replacement or supplementation of human interviewers with electronic bots could radically alter the cost structure and perhaps even quality of conducting surveys that depend on some direct respondent-

interviewer interaction. At a minimum it creates new possibilities for combining self-administration with personal interactions in mixed-mode surveys.

4. CONTACTING RESPONDENTS

I have already alluded to some of the ways in which new technologies can affect the way in which respondents are contacted. Basically there has to be some way in which potential respondents can be uniquely contacted. The two obvious methods are the telephone, whether by conventional telephones, cell phones or fax, and email. The absence of a method for randomly generating a list of email addresses precludes initial contact by email unless there is a preexisting list of addresses. I should note that traditional mail, now often referred to as snail-mail, suffers from a similar problem. While we have pretty good lists of postal addresses, we do not know who, if anyone lives at the addresses. They are of limited value for probability samples of individuals without some prior knowledge of who lives where, to say nothing of the response rate problems. Recently NSF was able to conduct a survey of grant recipients entirely by electronic means from advance letter to data file, but this was possible only because they had an up-to-date list of email addresses for the grantees. For the foreseeable future, however, there will have to be a separation between the initial contact mode and the mode used to deliver the survey itself.

The spread of wireless technologies may alter the situation by creating a world in which everyone is contactable wherever they are. Many of the contacting problems that currently exist come about because we have to depend heavily on the respondent being present at the fixed location of a receiving instrument. Wireless communication devices allow respondents to be contacted as long as they have the device with them and activated. For this technology to make an impact on survey practice three conditions would have to be met: 1) it would have to become almost universal, so as to have adequate coverage; 2) individuals would have to have unique device numbers, so as to be able to do probability sampling; and 3) there would have to be some way for respondents to indicate that they are willing to receive messages, so as to solve the cooperation problem.

Another advanced technology on the distant horizon is nanoscale chips that are imbedded in individuals. At present, this seems to be thought of as a device that would contain records, essentially personal and medical information about an individual that would be electronically readable in case of an emergency. Such devices could help with the locating problem if they were trackable by a GIS.

So far, in principle, there does not appear to be any technological barrier to a more complex development of nanoscale communication devices that would somehow be attached or embedded in individuals so that they literally are never separated from their wireless communication devices. Even if such devices were developed, however, many ethical and social considerations would have to be confronted before they could ever be employed on any scale that would be of use to survey researchers.

Separating the mode of contact from the mode of survey administration adds cost and increases administrative complexity, particularly for large samples. It thus limits the value of web-based survey methods. I will venture a prediction here. Web-based surveys will not replace telephone surveys at the speed that telephone surveys replaced face-to-face interviews until the email address list problem is solved. This is not to say that they will not be immensely popular among those who are willing to do SLOPs. I expect the popularity of such surveys to grow until we have a repeat of something like the Literary Digest failure. For surveys with serious claims to validity, the web as a mode of surveying will be limited to list samples or the specially recruited panels like that of Knowledge Networks.

5. QUESTIONNAIRE

I shall use the term "questionnaire" here to refer to data collection instruments even though I hope and expect that the rather dull, text-based questionnaires we are familiar with will be transformed into livelier and multi-media instruments that will be more engaging to respondents and provoke researchers to change the way they think about the measurement of behavior and attitudes. We have enormous opportunities here to improve not only in the way

we present questions to respondents, but also to rethink fundamentally our measurement models for behavioral and attitudinal reporting.

Developments in computer-assisted surveys over the past two decades have contributed to the solution of a number of the limitations of traditional paper-and-pencil questionnaire. For example, CATI/CAPI solved the skip problem and reduced errors from this source to near zero. Data entry is direct as respondents answer questions, and rotating questions or response alternatives can control order effects. Audio-CASI increases confidentiality by enabling respondents to answer questions directly without interviewer intervention while overcoming literacy problems. In longitudinal studies, pre-loading the computer with data from previous interviews greatly enhances the accuracy of reporting by providing memory aids to respondents as well as reducing telescoping. All of these developments have substantially increased the quality of surveys. For the most part, however, surveys are still delivered by means of a human agent and are in a fixed format.

As was foreseen by those who envisioned the information utility, the ability to download and upload information from households fundamentally changes the possibilities for surveys in two important ways. First, it expands the type of information that can be transmitted given sufficient bandwidth. It eliminates one of the major limitations of telephone survey, namely the restriction to auditory stimuli and responses. Web-based survey instruments can combine audio not only with static visuals but also with animations, video footage and text. Response modes can be auditory, textual, point and click, light pen or graphic. If household records become routinely held in electronic form, those surveys that are enhanced by consulting records will be simplified so that respondents can easily call up records to get answers to factual questions. Only our imagination and creativity limit the possibilities. Of course, a great deal of methodological research will be needed before we understand how such radically different modes of presenting information and alternative response modes affect our measurements. I trust that work of this sort will get started well before it is practically possible to realize many of these benefits.

The second important way that web-based surveys can change surveys is by making it possible for respondents to interact at their own convenience with what is in effect an electronic interviewer. One of the main features of present day survey methods, except for postal and web-based surveys, is that they require the interviewers and the respondents to be in contact simultaneously. Finding a convenient time for both the interviewer and respondent to be in contact is increasingly difficult and contributes to the decrease in response rates, particularly in telephone surveys. Web-based surveys enable respondents to answer at their convenience. They can break off and resume when they want. They can take time to consult records and to reflect on questions. They can undertake the task in their own time frame rather than being dependent on the limited windows of time that interviewers dare to call.

For the most part, surveys depend on verbal interchange between respondents and interviewers or on self-administered questionnaires. But information is also collected through administrative records, through analysis of web usage records, or, in some shopping studies, through the scanning of bar codes. Recently, however, there have been considerable advances in sensor technology and computer networking which open up the possibility that behavioral data might be collected directly and unobtrusively through the use of sensors. Because of increasing concerns about safety and security, the spread of sensors, including continuous video imaging, is growing apace. Whether or how a vastly larger use of sensors to monitor public (or even private) behavior will affect survey practice, I cannot tell. But it is a development that we should be aware of and consider its implications for survey practice.

Decoupling interviewers and respondents is not an unalloyed benefit, however. Important concerns about respondent motivation arise when the respondent is not under the social pressure of the interviewer to complete the interview, as we know from completion rates in postal surveys. The web mode enables us to deliver the survey to respondents without having to find them at home, and thus eliminates one of the big sources of low completion rates. Finding ways to encourage respondents to complete survey instruments is an important challenge for web-based surveys, but there will be more varied means at the disposal of those conducting the survey, including easier means to tailor and deliver specific incentives that might motivate respondents. Still it will be one of main challenges to the successful use of the method.

6. DATA PROCESSING

One of the greatest advantages of computer-assisted survey technology is the ability to enter responses electronically directly as respondents answer questions. Present technology, however, is still largely constrained to entering numeric data. While natural language data can be entered, it requires typing either by respondents or interviewers with attendant errors both of omission and commission. This technological limitation has led to two complementary results. First, we have practically abandoned open-ended questions or recording respondent comments and qualifications to their answers. Second, we have neglected the development of computer-assisted analytic methods for dealing with free text. Since we do not collect much free text, we do not feel it necessary to develop methods to deal with it. Similarly, since we do not have good methods for analyzing natural language responses, we do not bother to collect them. We need much better methods of directly representing natural language in electronic form to build up databases that can be used to develop analytic methods. And we need much more research on language and on methods of analyzing free text to measure the same types of variables we now measure through closed-ended multiple choice items. There are some exciting developments going on in IT, linguistics, and cognitive science that are addressing these issues. Survey methodologists need to learn more about them and bring them to bear on our needs.

7. OTHER ISSUES

I would now like to turn to several other issues that are perennial problems for survey research and consider what, if any, implications new technologies will have for them. Perhaps the principal problem confronting us today is the apparent decline in the willingness of respondents to cooperate in surveys. I believe that this is partly a consequence of the technological developments in the past few decades that have reduced the cost of doing surveys by telephone and the attendant rise of telemarketing. There has always been a tension between marketing and surveys. When door-to-door sales representatives were competing with face-to-face survey interviewers, the probability of being in a sample survey was small. I remember in the early days of my career at NORC there were complaints of confusion between Encyclopedia Britannica salesmen who represented themselves as coming from the University of Chicago, which did in fact own the Encyclopedia at that time, and NORC interviewers who also came from the University of Chicago. But it was a rare event because surveys, as well as door-to-door sales efforts were expensive and not all that common. Today, however, both telemarketing and legitimate surveys have proliferated as technology and the deregulation of the telephone industry have greatly reduced the cost of doing surveys via the telephone. The result is that potential respondents, even those who might be willing to participate in surveys, particularly government statistical or scientific surveys, may not take the time to distinguish the purpose of a call. In addition, what might be called "counter-technologies," such as answering machines, caller ID, and other methods of screening calls, have been developed to provide gate-keeping mechanisms that make it more difficult to make contact with a sample member and explain what the survey is about.

Marketing pitches and SPAM are already growing on the Internet, well before surveys have established themselves as major players. In so far as Internet surveys reduce the marginal cost of increasing sample size, we can expect that the natural tendency to have as large sample sizes as one can afford will push surveys toward larger and larger, perhaps even gargantuan sample sizes, thus increasing the probability that households or individuals will be selected in a random sample. The saturation of households with marketing and survey requests will no doubt spark the development of further Internet-based "counter technologies". The challenge here is to find a technology that will be able to deliver a differentiated message to those selected into a sample so that they can make an informed refusal, rather than rejecting the survey request out of hand.

A second issue that is of great concern to survey researchers and respondents is the confidentiality of data. When computer-assisted surveys were first beginning to be used, many survey researchers were worried that respondents would view the use of a computer to record and store their responses as less confidential than interviewers recording their interviews on hard copy. The reality, of course, is that it is a better means of keeping the data confidential because there were safeguards against outsiders getting to the responses, and the response data are in fact kept in files separated from the questionnaire so that even if someone had the data file, it would not be interpretable without a separate file that contained the questionnaire. In fact, studies of respondents' reactions to computer-assisted

personal interviews indicated that they somehow understood this and did in fact perceive them as being able to protect confidentiality better than paper-and-pencil formats (Baker, et al, 1995).

The web may be a different matter, however. Stories of computer hacking, the use for marketing purposes of responses to surveys or of past web viewing and buying patterns, theft of credit card information, etc. have raised serious questions about the confidentiality and security of responses to web-based surveys and, especially, the security of public use data files. As challenges to confidentiality arise, technological fixes to these challenges are invented. It is not clear at this point the degree to which people will continue to believe in the confidentiality assurances of researchers (Mackie and Bradburn, 2000, Doyle et al, 2001).

In 1970 I thought that an increase in electronic communication would result in a decline in people's interest in privacy and confidentiality. In one of my less astute predictions, I wrote: "...the whole distinction between public and private might be considerably reduced and only the old-fashioned, the subversive and the eccentric will be interested in privacy." (Bradburn, 1970, p. 286). Instead we have seen an upsurge of concern about privacy and confidentiality as people have discovered what can be done when computers are capable of storing massive amounts of data and records of electronic transactions, including even the most trivial encounters on the web, are routinely retained and can be mined at relatively low cost. The issue, which already had high saliency because of sensitivities to the increasing availability of medical records in electronic form, has taken on a new urgency after September 11th because many of the procedures for surveillance and detection of potential terrorists will involve putting together massive data bases that increase the potential for governmental agencies to know much more about individuals than is presently the case. How can such uses of technology be reconciled with the protection of civil liberties? The answers to this question may have considerable repercussions for survey research as well as for society.

8. CONCLUSIONS

I have tried to sketch out some of the possibilities and challenges I see facing survey researchers in the light of technological developments that either exist now or are going to be available in the near future. As with the concept of the information utility, there can be a long lag between what is technically possible and what has been commercialized and is readily available. Telephone surveys were possible and conducted on a limited basis in the 1930s and 40s, but until the 1970s telephone coverage did not become complete enough for telephone surveys to meet scientifically acceptable standards. Even then, the high price of long distance service restrained the unbridled use of telephone surveys. The reduction of the price in long distance service after deregulation made it even more attractive as a method both for market surveys and for telemarketing itself.

We are not now at a point where the penetration of Internet access is sufficient to easily conduct broad population surveys, even if the sampling issues could be solved. So we are limited to list-based samples where those in the list have email and web access or to special panels where a probability sample of households and individuals has been recruited and supplied with some common device as a platform for web-based surveys. I believe that we are in a transition period as with the telephone in the last two-thirds of the 20th Century. The Internet will become the major means of communication in the future. I do not know when this transition will be completed. I cannot say whether the web as we know it today will be the technology that triumphs or what will be the means that will be used to connect households to the Internet. But it will come.

We should use this transition period wisely and invest in methodological research that will address the issues I have raised here today and experiment with new ways of constructing questionnaires and new methods for analyzing data, particularly kinds of data that we have not had the opportunity to collect in the past. In short, we should exploit the possibilities that exist before us and stretch our imaginations to include doing things that would have seemed impossible even 10 years ago, even if they will not be practical for many years, or maybe never in the form that we now envision them.

It is an incredibly exciting time. We should make the most of it.

REFERENCES

- Baker, R., Bradburn, N. and Johnson, R. (1995), "Computer-assisted personal interviewing: an experimental evaluation of data quality and costs". *The Journal of Official Statistics*, Statistics Sweden., 11 (4), pp. 394-415.
- Bradburn, N. M. (1970), "Survey research in public opinion polling with the information utility--Promises and problems", In H. Sackman & N. Nie (Eds.), *The information utility and social choice*, Montvale, NJ: EFIPS Press, pp. 275-286.
- Blyth (1997), "Developing a speech recognition application for survey research", In L.E. Lyberg, P.P. Biemer, M. Collins, E. deLeeuw, C. Dippo, N. Schwarz, and D. Trewin (Eds.), *Survey measurement and process quality*, New York: Wiley.
- Dillman, D. A. (2000), *Mail and internet surveys: The tailored design method*. New York: Wiley.
- Doyle, P., Lane, J. I., Theeuwes, J. M. and Zayatz, L. V.(Eds.) (2001), *Confidentiality, disclosure, and data access: Theory and practical applications for statistical agencies*. Amsterdam: North-Holland, Elsevier.
- Groves, R. M., Biemer, P. P., Lyberg, L.E., Massey, J. T., Nicholls, W. L. II and Waksberg, J. (Eds.) (1988), *Telephone survey methodology*. New York: Wiley.
- Mackie, C. and Bradburn, N. M. (eds.) (2000), "Improving Access to and Confidentiality of Research Data: Report of a Workshop". *Committee on National Statistics, National Research Council*, Washington D.C.: National Academy Press.
- Sudman, S. and Bradburn, N. M. (1974), *Response effects in surveys: A review and synthesis*. Chicago: Aldine.