As with other technologies we suspect the initial costs will be unlikely to dampen CSAQ's future. In fact we believe that CSAQ software development and enhancements will continue at a rapid pace, driven by advances in computer technology and the rising penetration of the Internet, perhaps spurred by new low-cost options affordable by most households and the possible integration of the PC, television, and telephone into a common household appliance. A future in which all households are linked by flexible, interactive communications is clearly feasible. In such a future CSAQ data collection would surely have an important role.

CHAPTER 21

Ten Years of Interviewing Without Interviewers: The Telepanel

Willem E. Saris
University of Amsterdam

In 1984 the European Society for Marketing and Opinion Research (ESOMAR) organized a conference under the title: "Are Interviewers Obsolete?" At that conference Clemens (1984) introduced the first completely automated system for survey research without interviewers. Respondents were asked to answer survey questions interactively displayed on terminals in their homes or offices which were linked via telephone modems to remote central-computers in a research agency. Clemens used the British Videotex system for this purpose. French research agencies later employed the Minitel terminals that FRANCE TELECOM distributed to telephone subscribers for similar survey uses (Gautier, 1995). In the Netherlands, Saris and De Pijper (1986) of the Sociometric Research Foundation (SRF) began development in 1984 of the first Telepanel for use in computer assisted panel research of a representative sample of Dutch households.

The British Videotex System did not prove economically viable, and survey uses of the French Minitel system have been limited by its screen, entry ergonomics, and by the uneven distribution of Minitel computers across the French population. The first Dutch Telepanel has been functioning successfully since 1986 and has the longest, continuous history of nationwide interviewing without interviewers. Four additional Telepanels have been established since that time. This chapter describes (1) the historical development and special features of the Telepanel data collection method, (2) the social requirements of its operation, (3) the nature and quality of its sample, (4) the quality of its survey data, and (5) possible future developments.

Computer Assisted Survey Information Collection, Edited by Mick P. Couper, Reginald P. Baker, Jelke Bethlehem, Cynthia Z. F. Clark, Jean Martin, William L. Nicholls II, and James M. O'Reilly.
ISBN 0-471-17848-9

21.1 BACKGROUND

21.1.1 Origins

Computer assisted panel research (CAPAR) using computers in respondents' homes became feasible for nationwide surveys at about the same time as computer assisted personal interviewing (CAPI). The first portable microcomputers with keyboards and screens suitable for CAPI came on the market in 1985 (see Couper and Nicholls, Chapter 1). About two years earlier, inexpensive modems appeared on the market, making it feasible for home computers to receive and return information from remote central computers. Computer hobbyists used modems to reach central bulletin boards to exchange information and computer programs without charge, and commercial ventures, such as Videotex, began to sell subscription services to provide requested information for a fee.

In 1984 Saris and De Pijper recognized that these new capabilities of home computers could be used for survey research. If potential respondents had computers, telephones, and modems, a questionnaire could be sent from a remote central computer through telephone lines. Respondents could then respond to the questions, and their answers could be transmitted back to the central computer by telephone. We called this system the "Telepanel" because it used the telephone for the communications and the television as the computer "monitor."

The SRF field tested CAPI and the Telepanel at about the same time. We found that the Telepanel had two major advantages: First, since we generally preferred respondents to answer in self-administered mode, CAPI interviewers would have little to do other than bring the computer to resondents' homes and instruct them in its use. Second, since the Telepanel did not need to hire and pay interviewers, it should be much less expensive than CAPI.

The primary problem in using home computers for data collection is coverage. Most (or at least many) potential respondents in populations of interest will not possess a computer. This was especially true in 1985 when the first Telepanel was tested, and it remains true in 1998 as this chapter is written. This coverage problem can be solved in three basic ways.

The first is to restrict use of this data collection method to those specialized populations who own computers and modems or who have access to a shared computer system, such as the college population studied by Kiesler and Sproull (1986). Since very few populations meet these criteria, research opportunities are limited.

A second possibility is to take advantage of hardware placed in homes for commercial purposes and use it for survey research. For example, French market and opinion research companies have made use of the Minitel system that FRANCE TELECOM provided to many French telephone subscribers as a substitute for telephone directories.

The third possibility is for the survey organization to provide the necessary computing equipment to a sample of households who agree to participate in the survey. Clemens (1984) was the first to try this approach by providing Videotex equipment to a sample of doctors to obtain their views on medical issues. The same strategy was independently developed by SRF for its first experimental Telepanel of 100 Dutch households chosen in 1985 (Saris and De Pijper, 1986).

In 1986 an expanded Telepanel of 1000 households representative of the Dutch population was selected by the Dutch Gallup Organization (NIPO) for its opinion and commercial research. This original panel continues to serve that function. In 1991 two additional Telepanels were established in the Netherlands by the Telepanel Foundation (STP). The first is a 2000 household sample representative of the Dutch population. The second is a 1000 household sample drawn from the highest Dutch income decile. Telepanels of at least 1000 households also have been established in Finland and Italy. Thus a total of five Telepanels are currently known to be active.

21.1.2 Standard Telepanel Approach

All Telepanels other than the Dutch high-income panel were designed to be representative of their national household populations. They are probability samples of households who have agreed to be members of a continuing panel, participating each week in a survey that takes about 20–30 minutes per person. All members of panel households are asked to participate because the weekly surveys may target different demographic subgroups of the population. A maximum of about 2300 persons may be reached in the Netherlands with a 1000 household panel. Telepanel households are not paid for their participation although they are reimbursed for related telephone charges and other costs. The primary benefit of their participation, is having the Telepanel computer placed in their homes which may be used for other purposes. Panel households are given games and other software, and they can also arrange additional options such as access to the Internet. Supplementary rewards, such as participation in lotteries, are offered for unusually long or difficult questionnaires.

Each weekend, when a household member turns on the computer and selects the interview option, a new questionnaire is automatically downloaded via the modem. Data collection begins on Saturday. The first household member can then answer the questions or postpone survey completion until later that weekend. The system prompts for other household members as needed. When all have answered, the system automatically dials the central computer, uploads the answer files, and clears the diskette or hard disk for the next week. The household telephone is free for other uses except during downloading and uploading. Households who have not responded by Monday are telephoned by a member of the Telepanel staff. Data collection ends on Wednesday; and the analysis files and response reports are automatically generated. The next round of research begins the following day.

21.1.3 Comparisons with Other Methods

The Telepanel method is a form of computer assisted self-interviewing (CASI) for households in which the questionnaires are delivered by an electronic mail system (EMS). It differs from the CASI methods described by Tourangeau and Smith (Chapter 22) and Turner et al. (Chapter 23) because respondents answer the Telepanel questions on their own at their own pace. An interviewer does not have to be present when the questions are answered. The Telepanel also differs from the computerized self-administered questionnaire (CSAQ) and Internet/Web methods described by Ramos, Sedivi, and Sweet (Chapter 20) and Clayton and Werking (Chapter 27) in two important and related ways. First, because the Telepanel provides households with computers and modems, it avoids the coverage biases that CSAQ and Internet/Web surveys face through loss of potential respondents who lack the necessary equipment. Second, the Telepanel conducts surveys of households and household members while CSAQ and Internet/Web surveys have most often been used for surveys of organizations or businesses.

Because Telepanel respondents are committed to weekly participation for an extended period of time, the system obviously lends itself to panel surveys in which the same respondents are asked questions on the same topics in repeated waves. The STP Telepanel has been used for panel studies on consumption, income, savings, victimization, political party preference, and the effects of advertising. It differs from the large household panels described by Brown, Hale, and Michaud (Chapter 10) because they typically collect data only once or twice a year, while the Telepanel members participate weekly. Since many Telepanel weeks are not scheduled for panel surveys, ad hoc or cross-sectional studies are frequently conducted as well. Cross-sectional studies have been conducted on such topics as life satisfaction, current opinions, life histories, reactions to government decisions, and use of time. The total amount of information collected is very large, and it is all obtained from the same respondents. Individual studies can omit demographic details on the respondents who are already known from prior weeks.

The Telepanel resembles TV panels and consumer panels in its frequency of data collection but covers a much broader range of topics and interests. TV and consumer panels tend to concentrate on only one aspect of life. The Telepanel does share with TV and consumer panels the problem of avoiding a heavy response burden.

21.2 SOCIAL REQUIREMENTS

The social requirements of the Telepanel are as important as its technical requirements. Its success depends both on its ability to recruit, retain, and encourage panel members and on the quality of the software which should allow people of all backgrounds to participate. These two sets of requirements converge in the designs of the software and fieldwork.

21.2.1 Automated Operation

The process of downloading and uploading information between the central computer and the household's PC has been fully automated. If the PC is switched on and the interview option selected, the computer automatically makes contact with the central computer and downloads the interviews. Similarly the program detects when all household members have completed their interviews and automatically makes the connection to upload the responses to the central computer.

The survey questions must be designed for easy self-administration. The various types of questions and response options are shown to the respondents in an initial session when the computer is installed. In addition all instructions are available to the respondents at all times on each interview screen or in easily accessible help screens.

21.2.2 Help Desk

When technical problems arise, respondents forget the procedures, or they feel the need for personal contact, they may telephone a help desk which is staffed during each data collection period. Under normal circumstances about ten queries are received per four-hour shift in a sample of 2000 households. Most involve hardware failures and missing interviews or diskettes. These queries are easily handled by a single staff member who can perform other tasks between calls.

The help desk also serves a crucial function in alerting research staff to questionnaire errors. Such errors prompt many telephone calls and must be corrected as soon as possible with an updated interview file installed on the central computer. To avoid such problems and the irregular staff hours and respondent irritation they produce, questionnaires are extensively tested in advance.

21.2.3 Respondent Comments

In the STP Telepanels, respondents have the opportunity to enter comments or remarks at the end of each question and more general remarks at the end of each session. Respondents may report that they did not understand a question or could not correct an answer. The comments also are used for more general information, such as to explain that the household will be on vacation the following week, that they will be moving to another residence, or that they no longer want to continue in the Telepanel.

It is important that participants be given the opportunity to make comments and that these comments are read and acted on. Comments play a key role in maintaining two-way communication between respondents and the survey organization.

21.2.4 New Study Announcements

The Telepanels differ in the way they inform panel members that a new questionnaire is ready to be answered. The most common approach is by asking panel households to participate on a weekly basis, starting on the same day each week. In that way most households learn that it is their responsibility to initiate contact with the central computer each week. All Telepanels except the Dutch high-income panel use this as their primary method.

Although, in principle, the same procedure could be used for monthly rather than weekly surveys, a monthly schedule has several disadvantages: A day of the month is more difficult to remember (and control from the office) than a day of the week, monthly surveys could result in lengthier fieldwork periods, and new surveys would have to be announced through the system. This would be effective only if respondents left their computers on all the time. Most households will not do that because of the electricity costs. Therefore a mailed announcement is used for monthly surveys.

21.2.5 Panel Management and Personal Contact

All panels need a system to monitor questionnaire completion and to encourage the cooperation of panel members. The Telepanel, like other computerized systems, can quickly record and summarize questionnaire returns to identify participants who have not yet replied. Although electronic reminders can be sent to late respondents, they will only be effective for households who use their computers frequently enough to receive prompting messages or leave them on all the time.

To maintain high response rates and continuing panel participation, STP has found that personal relationships are beneficial, if not essential, between panel households and survey staff. Regional managers are employed for this purpose. Each is responsible for about 200 sample households and receives a database containing general information about these households, their participation over six months or more, and their entered comments about planned vacations, illnesses, or other matters that could affect participation.

The regional managers play a key role in nonresponse follow-up. They begin a weekly routine on Monday by telephoning all households who did not answer the latest questionnaire and have not previously communicated an explanation. If there is at technical or logistic reason for nonresponse, the manager will try to solve it immediately. Otherwise, the household is asked to respond by Wednesday morning. If the household cannot respond by Wednesday when fieldwork terminates, the manager asks the household the reason, which is included in the response reports.

In the evenings the regional managers contact those households who have entered messages in the past weekend's questionnaire about problems with questions or other matters. Respondents become irritated and may terminate their participation if such messages are not answered. In total, regional

managers spend about six hours a week overseeing 200 households. With time, regional managers get to know all their respondents and the respondents get to know their regional managers. This makes participation less anonymous for the respondents and encourages their continued participation.

21.2.6 Work Schedule and Staffing

Weekly research requires an efficient team and a tight time schedule as the following typical schedule suggests:

Day	*Task*
Wednesday	Select people for the next study.
Thursday	Test the final questionnaire and install it on the central computer.
Friday	Test the cleaning tasks and questionnaires.
Saturday	Data collection with help desk.
Sunday	Data collection with help desk.
Monday	Identify nonrespondents; call to remind delinquent panel members; conduct extra data collection.
Tuesday	Conduct extra data collection.
Wednesday	Create data files and response reports; identify those who will no longer participate; and select substitutes.

If these tasks are not completed on time, the workload builds up leading to serious problems. These tasks are primarily the responsibility of three people: one responsible for the questionnaires, one for the regional manager system, and one as systems manager. All other tasks, such as preparing the questionnaires for special studies, testing the questionnaires, and completing detailed analyses are carried out by additional staff. Some tasks, such as solving technical problems, choosing replacement households, and the placing of computers in their homes, can be spread over a period longer than a week. However, to maintain a stable sample size, the replacement tasks must not lag far behind.

21.2.7 Panel Administration

All panel studies face problems of maintaining current addresses of their panel members and of learning the new addresses of those who have moved (see Brown, Hale, and Michaud, Chapter 10). The Telepanel can often circumvent this problem as it maintains weekly contact with survey households. Households often provide information about a coming change of address in their weekly comments, and even if they fail to do so, their absence is immediately detected and results in a call by the regional manager. Locating respondents who have moved is simple under these conditions and very few households are lost in this process.

At the same time the use of PC equipment in the Telepanel generates problems that other panel studies do not have. This is not simply a matter of re-installing the equipment for households who move. Any household may experience technical difficulties with the computer, modem, or monitor. The technical staff first attempts to correct such problems by telephone. If that fails, a member of the technical staff goes to the household to exchange the computer or to make other repairs. Furthermore, to maintain management control of the whole process, a current inventory of all equipment and its status is needed. This requires more work than is necessary in paper and pencil panels but is similar to procedures for interviewer-administered CAPI. Fortunately the number of broken or stolen Telepanel computers is very small.

21.3 SAMPLE QUALITY

The quality of survey data collection depends on the quality of its sample and the quality of its responses. This section examines sample quality while Section 21.4 examines data quality.

21.3.1 Sampling Objectives

The objectives of the Telepanel sample design are based on its use for both panel studies and for ad hoc cross-sectional surveys. The Telepanel has many advantages for cross-sectional research, some of which are that a new sample does not have to be chosen each time; the research can be rapidly implemented; and interviews can be shorter because much information is already available about panel members. To be suitable for both cross-sectional and panel studies, the sample must be and remain representative of the population from which it is drawn. The sample requires adjustment over time to compensate both for panel attrition and for changes in the population (Duncan and Kalton, 1987).

The quality of the final sample for any specific application depends on four components: the sample design, the initial response rate in recruiting sampled households to the panel, panel attrition over time and substitution for those losses, and the response of the panel members in the individual study.

21.3.2 Sample Design

The sample must be continuously updated to match the relevant population characteristics. This could be accomplished with a systematically rotating panel. But such a design would complicate the study of individual change. Another alternative is the combination of a fixed and rotating panel to facilitate both panel and cross-sectional studies. Various cohorts could rotate in and out of the sample as needed. Alternatively, substitutes could be found for individual households who cease to participate. This can be done in such

a way that the sample remains representative of the population for the most relevant characteristics. This last option was chosen for the Telepanel. The STP National Telepanel used a two-stage stratified cluster sample for the initial selection (Geldrop, 1993).

21.3.3 Initial Nonresponse

The cooperation of sampled households in the Telepanel is obtained in stages. The households are first asked to participate in a very short telephone interview that obtains basic demographic and other background data. The telephone interview concludes by asking appropriate households whether they are willing to participate in a longer interview. If they agree, the next interview takes place in the respondent's home. This second interview is a computer assisted self-interview (CASI) with an interviewer present (CASI-IP). The interviewer brings the computer to the respondent's home. One purpose of this interview is to demonstrate how simple and easy self-interviewing using a computer can be. Only after this proves a positive experience are households asked to join the panel.

Not all households who complete the second interview agree to join the panel. Very few give technical reasons for refusing, such as saying that the computer is difficult to work with. The most common reasons for refusal are "no time" or "no interest" as in other panels or surveys. The initial response rate in the STP Dutch National Telepanel (slightly over 30 percent) is comparable to that of the Statistics Netherlands' Family Expenditure Survey. We should note that other surveys in the Netherlands also have low response rates. We hypothesize that response rates for Telepanels in other countries may be comparable to those of other high burden surveys such as expenditure surveys. Households with children are more likely to participate in Telepanels than those with no children (Geldrop, 1993). Cooperation in the Telepanel is also lower in large cities and among persons under 35 or above 65 years of age.

A two-stage sampling procedure can be used to compensate for these nonresponse deviations. In the first stage, a representative sample of approximately 10,000 potential households is asked to participate in the panel. In the second stage, households are randomly selected from the pool of households who have agreed to participate with probability inversely proportional to their likelihood of cooperation (Geldrop, 1993). In this way the unequal chances of participation can be corrected while retaining a probabilistic selection procedure.

21.3.4 Attrition

All panels face the problem of attrition as respondent participation diminishes over time. The size of the attrition problem is strongly related to the frequency of data collection and the response burden associated with each questionnaire. With the fixed automation costs of the Telepanel, cost effectiveness requires frequent use, and all Telepanels face continuing problems of attrition.

Sikkel (1994) showed that within 100 weeks after the start of the STP National Telepanel, only about 50 percent of the original households continued to participate. The tasks the respondents were asked to perform were so burdensome that most households were willing to participate only for a limited time. Sikkel also found that the hazard rate, or probability of dropping out, was highest after a year, but attrition naturally depends on the type of research, the number of questions per study, and the frequency of the data collection. After five years of experience, the STP National Telepanel has lowered the response burden, and the dropout rate fell by about half.

Several demographic characteristics have a significant effect on the hazard rate (Oppenhuisen, 1994), such as:

- *Age*: Young people drop out sooner. Although older people are more difficult to recruit initially, when they cooperate, they remain loyal longer than younger people.
- *Family size*: Small families drop out sooner.
- *Occupational status*: The retired and unemployed drop out more slowly. This suggests that people busy with jobs drop out sooner.
- *Income*: The highest income group drops out first.

The equipment the research organization provides can also affect participation. Households given a monitor remain longer in the panel. A separate monitor means that the computer does not have to be connected to the household TV so that answering survey questions does not interfere with TV watching. Respondents also are more satisfied and remain in the panel longer if they use their own home computer rather than the relatively simple computer provided by the research organization. Since the standard Telepanel computer had no hard disk, the interview process was slower and more tedious.

Differential attrition will bias the panel over time if lost households are not replaced by substitutions, thus taking account of any deviations between the panel and the target population. This is a problem all continuing panels face, especially those with frequent data collection. The STP Telepanel has solved this problem by selecting substitute households from the pool of potential respondents with a probability weight that is a function of the stratum the household was drawn from, the size of the stratum in the sample, and the stratum size required to represent the target population.

By applying this procedure on a continuous basis, the distributions of the stratifying variables in the sample will be comparable to those in the population (Geldrop, 1993). This does not mean that the sample is unbiased for other variables. That depends on the strength of the relationship between the stratification variables and the variables of interest.

21.3.5 Nonresponse by Wave

The last component of sample quality is the cooperation rate for individual weekly surveys. This has proved surprisingly high across the various Telepanels. In the initial period of the first Telepanel, NIPO reported a response rate of 95 percent or more of those who were able to participate (Saris, 1988). The latter qualification is necessary because there is a fluctuating number of people who cannot participate due to technical problems, illness, or vacations. With time these exceptionally high response rates decline, but even after five years the STP National Telepanel still reports weekly response rates of around 85 percent for households able to participate. Those not participating may have forgotten to notify the Telepanel of an absence or may be committed to other activities on that particular weekend. Since these events are approximately random, higher response rates are found for multiple weeks. In a typical two-week period of the STP Telepanel, the combined response rate was 91.2 percent, and for a three-week period, it was 93.4 percent of panel households.

Because the weekly response rates are so high, the representativeness of the weekly samples depends mostly on the representativeness of the continuing panel from which the households are drawn. Thus a great amount of effort must be spent on construction and maintenance of the panel and on procedures to avoid attrition and nonresponse.

Some of the most important methods of minimizing attrition and nonresponse are basic elements of the fieldwork design previously mentioned. They include:

- Well-designed questionnaires that are clear, informative, and cover a variety of topics over time to hold the interest of diverse respondents.
- A system of regional managers to maintain personal relationships with respondents, expedite the solution of technical problems, answer respondent questions, and follow up on respondent comments.
- The provision of useful computer hardware and interesting software to panel households to motivate their continued participation.

In addition lotteries can be used as an added incentive for households who are not interested in the hardware or software. As requested by the panel members, the lotteries emphasize one or two large prizes (e.g., video camera, or tickets to a major sports event) rather than many small prizes.

A newsletter for panel members also seems to be appreciated (see Sudman and Ferber, 1979). Newsletters present the results of completed studies, include photos of lottery winners, and describe other Telepanel activities, emphasizing that the panel members are participating in an important but pleasant activity.

21.4 DATA QUALITY

This section summarizes evidence on the quality of Telepanel data drawn from methodological studies conducted over the last ten years, focusing on three types of studies in particular. The first compares the quality of Telepanel data

with that of other collection modes under a multitrait-multimethod design. The second compares the quality of Telepanel data with that of other collection modes based on independent samples of respondents for each mode. The third examines the use of Telepanel data from complex survey instruments based on internal checks made during data collection. The complexity of these instruments would make cross-mode comparisons prohibitively difficult and expensive. This section concludes with a discussion of the effect of the Telepanel's high response burden on underreporting in survey panels.

21.4.1 Multitrait-Multimethod Studies

Multitrait-multimethod (MTMM) designs are perhaps best known for assessing alternative formats of attitude questions, for example, 100-point scales, 10-point scales, or ordinal categories. The same sample of subjects (respondents) is asked to answer the same set of attitude questions (traits) in each of the three formats. Using MTMM models and assumptions, the test-retest reliability and "true value validity" of each format may be estimated from the study data. The reliability and validity components also may be combined into a summary measure of data quality by format. (See Andrews, 1984, and Saris and Andrews, 1991, for a full description of these methods.)

The same MTMM design can be applied to comparisons of alternative data collection modes, such as CAPI, CASI, and their paper and pencil equivalents, rather than to question formats. The key requirements are that the same questions are asked of the same subjects (respondents) in all the modes being compared. The quality (reliability * validity) of each mode may then be estimated based solely on the observed data and MTMM model assumptions. The quality coefficients normally vary between 0 and 1 and can be interpreted as standardized regression coefficients. The coefficient squared is the variance in the observed scores explained by the latent trait(s) of interest.

Several MTMM studies have compared the estimated quality of Telepanel data with that of paper and pencil (P&P) or other computer assisted data collection modes (see Saris and van Meurs, 1990). One study, which focused on satisfaction questions asked in 15 different regions in Europe (Scherpenzeel, 1995), reported the following quality coefficient (in parentheses) for four collection methods: Telepanel (0.90), P&P personal interviewing (0.88), mail questionnaire (0.83), and P&P telephone interviewing (0.81). Using graphical line scales, which were only available in the Telepanel, the Telepanel quality coefficient was even higher (0.95) (Scherpenzeel, 1995). In a similar study of satisfaction questions in the Netherlands, Scherpenzeel and Saris (1995) report quality coefficients of 0.85 for the Telepanel, 0.68 for CATI, and 0.61 for CASI-IP. Another large MTMM study in the Netherlands found Telepanel data quality coefficients for several opinion items generally exceeding those from CASI-IP, which in turn were higher than those from CATI (see Scherpenzeel and Saris, 1997).

The Telepanel does not always obtain the highest-quality coefficients in these MTMM studies. For respondent evaluations of the seriousness of crimes, the quality coefficients were CASI-IP (0.67), Telepanel (0.57), and CATI (0.56). The evaluation of crimes was a difficult task for the respondents, as shown by the generally lower-quality indexes. The presence of the interviewer and the use of show cards in the personal interviews led to better performance (Scherpenzeel, 1995).

21.4.2 Mode Comparisons with Independent Samples

In the MTMM studies the same survey questions were asked of the same respondents in each mode and estimates of quality were obtained from their combined results. In traditional mode comparison studies, an independent sample is used for each collection mode, and comparable estimates of quality are not directly available. One can only examine such mode differences at an aggregate level. With independent samples, additional sources of variation, such as differential response rates by mode, may confound the comparison as may specific procedures necessary to adapt the questions or study objectives to the specific mode. The results are less pure mode comparisons than system comparisons. Several such studies have nevertheless provided useful information about the data quality of Telepanels.

Kalfs (1993) compared estimates of media use reported in Telepanel and CATI surveys. She found that reported daily television watching by highly educated people was about 30 minutes higher in the Telepanel than in CATI. This difference seems attributable to the social desirability bias when the questions were asked by a CATI interviewer. Similarly Bon and van Doorn (1988) found that drug use was between 15 and 20 percent higher when reported in a Telepanel than in face to face interviews.

The special capabilities of computer assisted data collection might also make important contributions to mode effects. In comparative time budget studies, Kalfs (1993) found that 50 percent of the trips reported in Telepanel data collection were left unreported in paper and pencil diaries. In the Telepanel time budget study, a consistency edit looked for a trip report whenever a change of location was reported. Such interactive edits were not possible in P&P diaries.

21.4.3 Data Quality in Complex Questionnaires

The design of complex questionnaires for the Telepanel and other forms of CASI requires special care because respondents must answer on their own. Typically no interviewer is present to guide the process, correct errors, and provide help when needed. All assistance must be built into the interview program. The screens should be simple, the instructions clear, and help screens readily accessible.

Assistance with navigation, or movement from screen to screen, is especially important when no interviewer is present. One important means of avoiding navigation problems in CASI is the use of summary and correction (S&C) screens (Saris, 1991). Respondents inevitably make occasional mistakes in answering survey questions, and if those mistakes occur in questions later used for routing (or branching), the consequences can be serious. If the erroneous routing paths lead to requests for information about income sources mistakenly chosen or persons no longer in the household, the respondent may be sufficiently annoyed to terminate the self-interview. S&C screens located at strategic points throughout the instrument summarize data already entered and ask the respondent to verify the information and make any necessary corrections.

In practice, Telepanel respondents frequently make corrections in the S&C screens and thereby avoid many potential errors. The S&C approach is more efficient than asking respondents to back up in the questionnaire to the original entry and make the correction there, something that even experienced CAI interviewers often find difficult (see van Bastelaer, Kerssemakers, and Sikkel, 1988).

The effectiveness of S&C screens has been demonstrated in constructing calendars in life history research. STP Telepanel respondents were asked to list in chronological order major events in various domains of their lives and to provide the dates for each of those events. The Telepanel program then constructed a S&C calendar summarizing these events and dates. The respondents were then asked to make any necessary corrections. About a fifth (21%) of the respondents corrected the order or events and almost four-fifths (79%) corrected the dates after seeing the information summarized in a single calendar (Vis and Wouters, 1996). Without a S&C screen these errors would have remained undetected. As shown by Holt, McDonald, and Skinner (1991), errors in dates can have especially devastating effects on parameter estimates in event history models.

The data summarized on S&C screens may include information from prior panel waves as well as from the current one. In the Telepanel, past information can be efficiently stored on the household computer's hard disk for use in later waves. Data from prior waves are especially important for income surveys where they can be used for sophisticated edits in dependent interviewing.

Verwey et al. (1989) has described a study in which the reported income of the household head fluctuated substantially from month to month. The median income of household heads in the Netherlands at that time was about 2200 Guilders (NLGs) per month (NLG 1.60 $\approx$ 1 U.S. dollar at the time). Monthly fluctuations of more than NLG 150 were found for 12.7 percent of the respondents, while 5.6 percent had monthly changes of more than NLG 500. In an initial study no special methods were used to check the quality of the data because such large fluctuations were not anticipated. In a subsequent study edits were added using the income reported in the prior wave.

If a respondent reported a monthly income NLG 50 greater than in the previous month, he or she was asked whether the reported income was correct and shown the previously reported amount on the same screen. This probe or check question was asked only under the specific condition that

$$I(t) > I(t-1) + 50 \quad \text{or} \quad I(t) < I(t-1) - 50$$

where $I(t)$ is the income just reported and $t-1$ is the income reported on the previous wave. The screen appears as follows:

> Perhaps you made a mistake because last time you reported that your income was "$I(t-1)$." If the amount was incorrect, return to the previous question by pressing the F1 key, otherwise press ENTER.

Brown, Hale, and Michaud (Chapter 10) have called this the "reactive" approach to dependent interviewing. In this case the check was not very effective because the month-to-month fluctuations in reported income were not reduced (Verwey, 1992).

Since we were still not convinced that such large monthly income changes were possible, a third study added another procedure called "scheduling" (Saris, 1991). Before the current month's income was asked, last month's reported income from 23 possible income sources was displayed on the screen, and the respondent was asked to correct any of those figures that were wrong. In this way an effort was made to prevent as many typing errors as possible while allowing for large changes to be reported (Hartman et al., 1991). This approach is called "proactive" by Brown, Hale, and Michaud (Chapter 10). The question appeared as follows:

> Our records indicated that your income last month from {Source S} was "$I(t-1)$." If this is correct, press ENTER. If this amount is incorrect, type the proper amount here _____.

Then after the current month's income was entered, the reactive check described above also was used. Scheduling and checking were used consecutively. The results of checking alone and of scheduling with checking on reported income of the household head are shown in Table 21.1.

The result of these scheduling and checking procedures was that minor typing errors were reduced and reported monthly income fluctuations decreased. The revised estimate of monthly income fluctuations may have been partly an artifact of the procedures used. We accepted this possible bias to see whether the large monthly income changes would disappear. Although the monthly changes became smaller, rather large fluctuations still remained, as can be seen in Table 21.1. Once all possible checks were built in and major monthly changes remained, we concluded that large fluctuations in monthly income do occur (Saris, 1995).

Table 21.1. Changes per Month in Income for the Head of the Household (in %)

	Checks in Data Collection		
	No Checks	Reactive	Proactive and Reactive
Change in NLG			
Over 500	5.6	9.9	3.3
150–500	7.1	12.6	4.8
50–150	10.1	6.9	6.6
5–50	13.6	6.9	5.7
0–5	61.6	63.7	79.7
(Sample size)	(950)	(950)	(1500)

Similar procedures have been developed to check the quality of prices reported in family expenditure studies (Hartman et al., 1991; Saris, Prastacos, and Marti-Recober, 1995). In these studies respondents are asked how much they had purchased of various commodities and the total amount they paid. For most goods the national average price per unit is known and can be stored in the instrument as a check for the price reported by the respondent. This is somewhat inefficient because prices vary rapidly over time and any given household can pay a variety of prices for a single good that they purchase on a regular basis.

To address these concerns, a procedure was developed that takes into account both variations in the families' buying patterns and the price changes those families experience over time. This is done by using the following condition to evoke the edit check question:

$$P(t) > P(t-1) + a(t-1) \quad \text{or} \quad P(t) < P(t-1) - a(t-1)$$

where $a(t-1) = [a(t-2) + \{P(t-1) - P(t-2)\}]/2$, $P(t)$ is the price per unit calculated from the price specified in the interview, $t-1$ is the time of the previous interview, and $t-2$ is the time of the interview before the previous one.

The use of $P(t-1)$ in the check makes this test family specific by using the last price this family paid. The use of the interval $a(t-1)$ makes the edit dynamic, since the interval is a function of fluctuations in the price *this family paid* over the two prior time periods. When past price fluctuations have been large, so is the interval. If the price does not fluctuate from wave to wave, the interval is halved for each interview until a minimum is reached. Too narrow an interval could annoy respondents. When the condition for the edit check is satisfied, participants see the following screen:

> You may have made a mistake. We have recorded that a (kilo/liter) of {goods} costs NLG "$P(t-1)$" while you paid NLG "$P(t)$." Could you please indicate which of the following was the case?
> 1. This is now the normal price.
> 2. The price is correct but I got a special offer.
> 3. The price was correct but this time I paid more.
> 4. The price entered was incorrect.
> 5. The amount entered was incorrect.
> 6. Both were incorrect.

If the respondent indicates that the wrong price or amount was entered, he/she is asked to correct it. If the entry is correct, the respondent is asked to indicate whether the entered price is now the normal price. If it is, the new price is stored for further use, and the interval $a(t-1)$ is adjusted on the next wave. If the respondent indicates that the family paid more or less this time only, no change is made to the price database.

The stored information can be used to impute missing information as well as for consistency checks. For example, if the family reports the total price of purchased goods but not the amount, the computer can suggest the amount, as in the following screen:

> Given the price per kilo/liter and the total price you paid, we think that you bought {*estimate*} kilos/liters. Do you think that this is correct?
> 1. Yes
> 2. No
> 3. Don't know.

If the respondent answers "yes," the estimate provided is recorded as the answer. If the respondent answers "no" or "don't know," a missing value is recorded. Total costs can be similarly imputed and suggested from amounts entered. In this way many missing values can be avoided. And since the suggested amount or total cost is based on the family's most recently reported purchases of those goods, it is more useful and accurate than a missing value.

With such procedures to ensure data consistency and impute missing values during self-interviewing, Maartens (1995) concluded that it was no longer necessary to clean the data after collection. Postinterview corrections led only to nonsignificant and substantially irrelevant changes in consumption estimates at the aggregate level. This clearly shows the benefits of dependent interviewing and dynamic range checks as employed in the Telepanel. In family budget research, postinterview data editing typically takes a lot of time and effort and still may not provide better data than obtained directly from the Telepanel.

This does not mean that the Telepanel data are without error. Respondents will at times report very deviant total costs or other quantities and fail to correct them in summary and correction screens. The Telepanel also is subject to other biases that are common to all panel surveys that collect data frequently.

21.4.4 Limitations of the Telepanel

Panels that ask respondents for a significant portion of their time each week create a high response burden which may have unfortunate consequences. In addition some studies, such as consumer research, may be seen as more burdensome than others. Consumer research requires a major effort from respondents; the topic is less interesting than social or political subjects, and these studies often are highly repetitive. If the purchase of consumer goods is the topic too frequently, panel attrition will increase.

Underreporting can be an even more serious consequence of high response burden in panel studies. The effect of panel participation on underreporting has

Table 21.2. Recorded Amounts of Selected Goods over a Period of Six Months Expressed Relative to the Base Month (Base = 100)

	Month					
Goods	Jan	Feb	Mar	Apr	May	Jun
Red meat						
Expense	100.0	99.3	91.0	93.8	89.7	88.9
Volume	100.0	95.4	87.4	90.8	82.8	85.1
Number of products	100.0	99.4	92.9	92.9	89.6	90.3
Meat products						
Expense	100.0	101.0	92.8	91.0	90.6	82.3
Volume	100.0	103.0	90.9	84.8	84.8	78.8
Number of products	100.0	102.1	95.9	92.8	87.2	84.1
Poultry						
Expense	100.0	87.5	82.1	80.2	79.0	75.5
Volume	100.0	80.8	76.9	73.1	69.2	76.9
Number of products	100.0	90.9	87.9	81.8	81.8	75.8
Eggs						
Expense	100.0	91.8	86.6	90.7	73.2	70.1
Volume	100.0	91.3	85.4	92.1	75.7	75.7
Number of products	100.0	91.3	85.4	92.1	75.7	75.7
Total						
Expense	100.0	98.0	90.2	91.1	87.8	84.5
Volume (excluding eggs)	100.0	93.8	87.0	87.0	80.8	82.2
Number of products	100.0	99.8	93.3	91.3	86.7	84.8

Source: STP panel, $n = 1750$.

been treated in many studies (Kemsley, 1961; Bailar, 1975; Olivier, 1987; Silberstein and Scott, 1991; van Meurs, van Wissen, and Visser, 1989). The Telepanel provides an additional example.

Table 21.2 summarizes reported purchases of a variety of products over a six-month period in 1994 in the STP National Telepanel (Kaper and Saris, 1996). Total purchases of all four commodities decreased substantially over time, except in April which included Easter when the Dutch traditionally purchase and consume more eggs. It is implausible that the downward trend represents a true decline in consumption; these figures were collected before the onset of mad cow disease caused an actual decline in meat consumption. It is also not likely that these effects are the result of telescoping. The respondents were asked to report their purchases weekly. The only obvious explanation for this decrease is respondent satisficing as a consequence of this study's heavy response burden. It is also clear that in this form the resulting estimates cannot be used for any practical purpose.

Underreporting due to response burden is likely to occur in any panel with frequent measurement and requiring active respondent participation. There are several possible solutions to the underreporting problem.

One option is to screen the respondents more carefully, only retaining those who seem to respond in a serious and stable way. The difficulty with this strategy is that people retained in the panel may no longer be representative of the total study population. In one study of this approach, Oliver (1987) found that those retained in the panel disproportionately reported the purchase of inexpensive products.

A second solution is to estimate the effect of nonresponse, refusals, and underreporting on the aggregated data and to adjust for these factors (Van den Oord and Saris, 1994; Kaper and Saris, 1996). This procedure requires extensive knowledge of the interrelationships among the variables of interest, which can only be developed over an extended period of time.

A third solution is to attempt to control the data quality by personally contacting each household when it reports deviant results. This strategy is relatively expensive in staff time and requires a staff with exceptional ability, dedication, and continuity. If the staff efforts in contacting deviant households vary, this may cause fluctuations in the data that have nothing to do with real change.

A fourth approach to controlling data quality is through the use of the proactive and reactive edit checks described above. For example, if a given household normally buys meat but then stops reporting meat purchases, an additional question can ask why meat was not bought. If the probes are sufficiently elaborate, omitting a purchase will no longer reduce the interview time, eliminating a motivation for underreporting. This strategy has been investigated by van den Oord and Saris (1994).

In view of the costs or limitations of all four alternatives, it may be better to attack the problem directly by reducing the response burden. Hogendoorn and Sikkel (1994) completed a statistical analysis that indicated that measuring

consumer purchases every two weeks, instead of every week, would reduce underreporting. A reduction in the total response burden of the STP household Telepanel also seemed to alleviate attrition rates. When the total response burden was reduced by approximately half, the attrition also dropped by half. Thus the best solution to the problems associated with a high response burden seems to be a reduction of that burden.

21.5 EVALUATION AND DISCUSSION

This chapter summarizes 10 years of Telepanel development and research. It identifies several favorable features of the Telepanel, including the high weekly response rates and the quality of its data in comparison with other data collection modes. These data quality benefits may not be realized unless study directors invest the resources to develop the necessary edit checks. The effort required to design and test data quality improvements is most cost effective in repeatedly used questionnaires (Saris, 1991). The high quality of Telepanel data partially stems from its use for panel and repetitive cross-sectional surveys where data quality improvements are more easily supported.

This chapter also identifies a major limitation of the Telepanel. Its high response burden can result in attrition and the underreporting of behaviors measured frequently, especially in consumer research. These problems are common to all panels that collect frequent observations on the same topic, unless their data collection procedures are fully automated and make little or no demands on the respondents' time.

In view of these common problems, an appropriate future direction for survey methods should include the further automation of data collection for frequently measured factual behaviors. This has already happened in TV audience research where the "People Meter" almost completely automates the measurement of household TV viewing (Buck, 1987; Saris, 1989). The use of bar code scanners to record purchases, both in stores and in respondents' homes, has brought high levels of automation to consumption and expenditure surveys. Similar procedures, possibly involving the use of smart cards, may in time be extended to the measurement of doctor and dental visits, hospital stays, and other recurrent behaviors.

The need to ask respondents questions about the reasons for their behaviors and about their plans, intentions, life satisfaction, attitudes, and opinions will remain. For these purposes the Telepanel and other forms of CAI will have a major role, although they may take new forms. The experience of the Telepanel suggests some limits on these future systems.

Many writers on survey methods have suggested that the Internet will be the primary medium of future data collection. Whether this happens depends on whether access to the Internet will be free or will provide other general benefits the public is willing to pay for. Thus far all data collection systems that households have had to pay for have failed to achieve a sufficient penetration

of the population to yield a representative sample of households. This applies to the Videotex system and to all tests to date of paid Interactive TV. But, if at some point Internet access is made possible at minimal cost through a TV or other common household appliance, the Internet may expand very rapidly as a vehicle for survey research. Should that happen, research organizations will be able to undertake interviewing without interviewers without making large investments. If the Internet does not achieve broad household penetration, the Telepanel system will remain an attractive option for continuous research on representative samples of the population.

In either event, the general kind of system this chapter describes may become the standard model of opinion research in the future, whether that system functions through the Internet, through computers placed in respondents' homes, or in other ways. The requirements for frequent survey research based on a continuing panel will still apply, including largely automated operation, simple, clear questions, help desks, regional managers to foster two-way communication, systems to announce new studies, continuous panel management, and so on. The procedures for data quality improvement discussed in Section 21.4 also could be used under a variety of possible systems, but to be effective, they will require as large an investment in time as they did in the Telepanel.

ACKNOWLEDGMENTS

I am grateful for the useful comments on an earlier draft of this chapter from the following colleagues: Robert Voogt and Corrie Vis of the STP Telepanel; William Nicholls, Magdalena Ramos, Barbara Sweet, and Elisabeth Sweet of the U.S. Census Bureau; Richard Clayton of the U.S. Bureau of Labor Statistics; and Roger Tourangeau and Tom Smith of the National Opinion Research Center.