Produced by
Section on Survey Research Methods
American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415 USA

ASA Series
What Is a Survey?

How to Collect Survey Data

Section on Survey Research Methods
American Statistical Association

This pamphlet, **How to Collect Survey Data,** is the third in ASA's newly revised series **What Is a Survey?** It discusses the actual steps in collecting survey data. Examples are drawn primarily from household samples; the emphasis is on operational issues and recent changes arising through survey automation.

The **What is a Survey?** series is written primarily for the general public. Its overall goal is to improve survey literacy among individuals who participate in surveys or use survey results. The series is designed to promote a better understanding of what is involved in carrying out sample surveys—especially those aspects that have to be taken into account in evaluating the results of surveys.

How to Collect Survey Data

Earlier in this series, we briefly discussed the main types of surveys—**mail, telephone interview,** and **in-person interview.** Each form has distinct advantages and disadvantages. Each is also greatly being changed by the use of computers and telecommunications. This pamphlet discusses current practice and the innovations now in process for all three.

How to Collect Self-Reported Data

Currently, mail surveys are the most common example of self-reported data collection. One reason is that these surveys can be relatively low in cost. This does not mean, however, they are necessarily easy to carry out. Planning the questionnaires for mail surveys is often more difficult than for surveys that use interviewers. F*or example,* care is needed to anticipate issues that respondents may have and to deal with them ahead of time.

Mail surveys are the most common example of self–reported data collection.

Using the mail can be particularly effective in business surveys—such as those conducted by the U.S. Bureau of the Census or the U.S.

Bureau of Labor Statistics. Mail surveys also work well when they are directed toward specific groups—such as, subscribers to a specialized magazine or members of a professional organization.

The manner in which self-reported data are obtained has begun to move away from the traditional **mail-out/mail-back** approach. The use of fax machines—and now the Internet—is on the rise. Fax numbers and Internet addresses are being added to specialized membership and other lists. As a by-product, they can be used, along with more conventional items like names and mailing addresses, in building potential sampling frames.

There are still other methods of obtaining self-reported data. F*or example*, the U. S. Bureau of Labor Statistics has a panel of business establishments, in which the respondents supply monthly data via touch-tone telephone entries that are directly connected to the agency's computers.

For the immediate future, this type of automation will probably be restricted largely to business or institutional surveys in which the same information is collected at periodic intervals—monthly, quarterly, etc.

As computers and telecommunications become more widespread, touch-tone applications or those involving respondents' computers *"talking"* directly to the survey organizations' computers will increase significantly. This increase is already well underway in health surveys, where samples of patient records are often supplied electronically.

How to Conduct a Survey Interview

Interview surveys—whether face-to-face or by telephone—offer distinct advantages over self-reported data collection. The *"presence"* of an interviewer can increase cooperation rates and make it possible for respondents to get immediate clarifications.

The main requirement for good interviewers is an ability to approach strangers in person or on the telephone and persuade them to participate in the survey. Once a respondent's cooperation is acquired, the interviewers must maintain it, while collecting the needed data—data that must be obtained in exact accordance with instructions.

For high-quality data to be collected, interviewers must be carefully trained through classroom instruction, self-study, or both. Good interviewer techniques are stressed, such as...how to make initial contacts... how to conduct

interviews in a professional manner...and how to avoid influencing or biasing responses. Training generally involves practice interviews to familiarize the interviewers with the variety of situations they are likely to encounter.

Time must be spent going over survey concepts, definitions, and procedures. A question-by-question approach is needed to be sure the interviewers can deal with any misunderstandings that may arise.

In most reputable survey organizations, the interviewers are also required to take a strict **oath of confidentiality** before beginning work.

Survey materials must be prepared and issued to the interviewers. For traditional paper-and-pencil, in-person interviews, ample copies of the questionnaire, plus a reference manual, information about the identification and location of the households, and any cards or pictures to be shown to the respondents must be given to the interviewers.

Before conducting in-person interviews, survey organizations frequently send an advance letter to the sample respondents, explaining the purpose of the survey and that an interviewer will be calling soon.

In many surveys, especially those sponsored by the federal government, information must be given to the respondents regarding the voluntary or mandatory nature of the survey and how the answers are to be used.

Visits to sample units are scheduled with attention to such considerations as the best time of day to call or visit, and allowance is made for repeated attempts (*i.e.*, *callbacks*) in not-at-home situations.

What Is CATI?

The use of computers in survey interviewing is becoming quite common. In the United States, most of the large-scale telephone surveys are now conducted via **CATI (*Computer-Assisted Telephone Interviews*).** With CATI, the interviewers use a computer terminal. The questions to be asked appear on the computer screen, and the interviewers use the keyboard to directly enter the respondents' replies as they are given.

CATI's important advantages are in *quality and speed,* not in cost savings. CATI can cost more for small, non-repeated surveys, due to programming the questionnaire. CATI's cost per interview decreases as sample size increases—so in large and/or repeated surveys, it is cost competitive with conventional telephone methods.

The use of computers in survey interviewing is becoming quite common.

The CATI interviewer's screen is programmed to show questions in a planned order, so that interviewers cannot inadvertently omit questions or ask them out of sequence. F*or example*, the answers to some questions require "***branching***" (*i.e.*, *answers to prior questions determine which other questions are asked*). CATI can be programmed to do the correct branching automatically. In noncomputer-assisted telephone interviewing, incorrect branching has sometimes been an important source of errors, *especially omissions.*

In the CATI setting, the computer can be programmed to edit replies. The computer can check whether the reply to a particular question is consistent with other information reported. If the editing indicates that a prob-

The latest poll indicates we can still avoid doing anything for the next quarter if we keep taking polls...

POLL DU JOUR
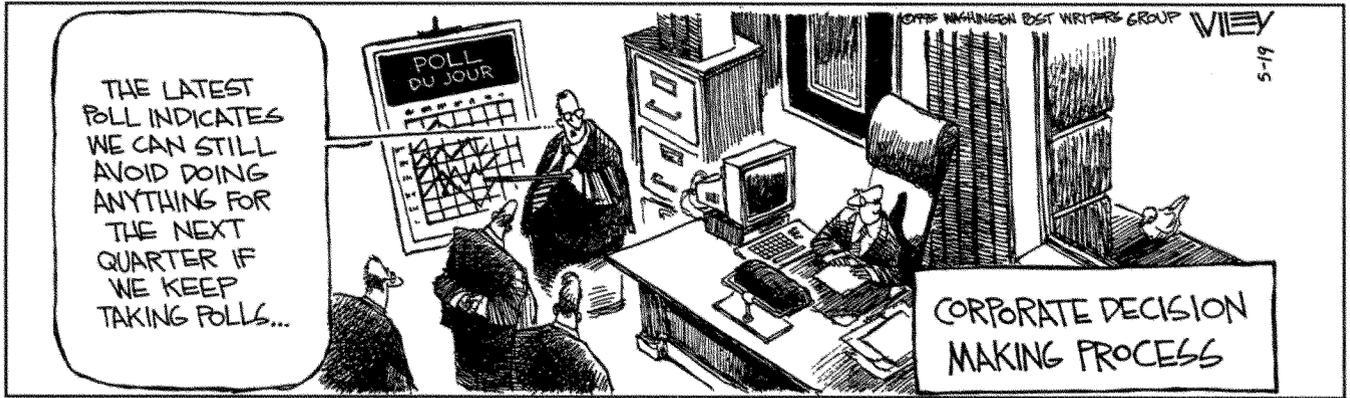
©1995 Washington Post Writers Group   WILEY   5-19

CORPORATE DECISION MAKING PROCESS

lem may exist, the respondents are asked to confirm or correct earlier answers.

CATI can produce statistical results quicker than traditional methods of data collection. F*or example*, it eliminates the need for a separate computer data-entry step. Furthermore, with CATI, some organizations are able to provide summaries of results as each questionnaire is completed or at the end of each day,

What About CAPI?

In recent years, there has been a trend toward the use of portable laptop computers for in-person interview surveys.

Portable computers can be taken into the field, and either the interviewer or the respondent can directly enter data in response to questions. Data collection carried out in this way is referred to as **CAPI (*Computer-Assisted Personal Interviews*).**

The CAPI laptops are not directly connected with a centralized computer. Nonetheless, most CATI quality and speed advantages also occur with CAPI.

Although only a few organizations currently employ CAPI methods, their use is expected to expand in the next few years, F*or example*, the

very large monthly Current Population Survey, which measures unemployment, has recently been converted from conventional in-person and telephone interviews to a combination of CAPI and CATI.

Clearly, as electronic technology becomes more widely used, traditional paper and pencil methods may eventually disappear—at least in surveys conducted by the federal government.

...there has been a trend toward portable laptop computers for in-person interview surveys.

What Is Done After Data Collection?

No matter what type of data collection is used, there are a number of **"*back-end*"** processes that may be needed to get the data in a form so that aggregated totals, averages, or other statistics can be computed.

For mail surveys and conventional paper and pencil interviews, this may involve coding after the questionnaires have been completed. Coded paper questionnaires are entered into a computer (*e.g., being keyed onto a disk*) so

that a computer file can be created. At this point, most of the remaining back-end steps are common to all surveys, whether or not a computer was used initially for data collection.

Once a computer file has been generated, additional computer editing, separate from clerical editing, can be accomplished to alter inconsistent or impossible entries.

No matter what type of data collection is used, there are a number of "back–end" processing steps.

Decisions are usually needed on how to handle missing items—cases in which the respondent did not know the answer... refused to provide one...or in which the question was simply not asked. Preferred practice for missing items is to provide special codes indicating why the data are not included. When resources are available, the "*filling in*" or imputation of these missing data items should be undertaken to reduce any biases arising from their absence.

When there is a "*clean*" file the survey data are ready for analysts to begin summarizing what has been learned. It is a good idea to use commercially available software packages to carry out this step rather than using your own specially written computer programs.

Often the best way to start the analysis is with simple counts and related percentages for each question. Next, it is common to produce tables of growing complexity. Eventually, there may be a need for even more sophisticated forms of data presentation to address the concerns outlined when the survey was initially conceived.

The results of surveys are usually printed in publications and presented at staff briefings or in more formal settings. Additional analyses can also be done by making unidentifiable computer data files available to other researchers at a nominal cost.

Shortcuts to Avoid

Conducting a credible survey entails scores of activities, each of which must be carefully planned and controlled. Taking shortcuts can invalidate the results and badly mislead the sponsor and other users. Here are three shortcuts to avoid, that crop up often:

- Not pretesting field procedures
- Not sufficiently following up on non-respondents
- Sloppy fieldwork and inadequate quality controls.

A **pretest of the questionnaire and field procedures** is the only

A pretest of the questionnaire and field procedures is the only way of finding out if everything "works."

way of finding out if everything "*works*"—especially if a survey employs new techniques or a new set of questions. Because it is rarely possible to foresee all the potential misunderstandings or biasing effects of different questions

and procedures, it is vital for a well-designed survey operation to include provision for a pretest. There should usually be a series of small-scale pilot studies to test the feasibility of the individual techniques (if new) or to perfect the questionnaire concepts and wording.

This should be followed by a full-scale "*dress rehearsal*" to find out if everything connects together as intended.

Failure to **follow up nonrespondents** can ruin an otherwise well-designed survey, It is not uncommon for the initial response rate in many surveys to be under 50 percent.

To deal with this possibility, survey plans should include returning to sample households where no one was home (*perhaps at a different time or on a weekend*), attempting to persuade persons who are inclined to refuse, and so on. In the case of mail surveys, it is usually necessary to conduct several follow-up mailings—spaced, possibly, about three weeks apart.

There is some evidence that responses to subsequent mailings may differ from responses to the first mailing. Thus, biases can result without the extra effort of follow-ups. Depending on the circumstances, it may even be necessary to contact a subsample of the remaining nonrespondents by telephone or personal visit.

A low response rate does more damage in rendering a survey's results questionable than a small sample, because there may be no valid way of scientifically inferring the characteristics of the population represented by the nonrespondents.

Sloppy execution of a survey in the field can seriously damage results,

Controlling the quality of the fieldwork is done in several ways, most often through observation or redoing a small sample of interviews by supervisory or senior personnel. There should be at least some questionnaire-by-questionnaire checking, while the survey is being carried out; this is essential if omissions or other obvious mistakes in the data are to be uncovered before it is too late to fix them.

In other words, to assure that the proper execution of a survey corresponds to its design, every facet of a survey must be looked at during implementation. F*or example*... reexamining the sample selection...redoing some of the interviews... assessing the editing and coding of the responses.

Without proper checking, errors may go undetected. With good procedures, on the other hand, they might even have been prevented. Insisting on high standards in recruiting and training of interviewers is crucial to conducting a quality survey,

Just looking at each step by itself is still not enough. As **W. Edwards Deming recommends,** a complete systems approach should be developed to be sure each step fits into the previous and subsequent steps. Murphy's Law applies here, as elsewhere in life. The corollary to keep in mind is that not only is it true that "If anything can go wrong it will… but, "If you didn't check on it, **it did.**"

Where Can I Get More Information?

In addition to the pamphlets in this series, ASA also makes other brochures available upon request:

■ **Ethical Guidelines for Statistical Practice**

■ **Surveys and Privacy,** produced by the ASA Committee on Privacy and Confidentiality.

For the above brochures or other pamphlets in the **What Is a Survey?** series, contact:

Section on Survey Research Methods
American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415 USA
(703) 684-1221/fax: (703) 684-2037
Email: asainfo@amstat.org
Web site: http://www.amstat.org/
sections/srms/

Besides the ASA, there are many other associations that are concerned with the proper collection and use of survey data:

■ **The American Association for Public Opinion Research** (AAPOR) offers a number of publications—perhaps the most relevant of these is the one entitled **Best Practices for Survey** and **Public Opinion Research Survey Practices AAPOR Condemns.** To obtain copies, call (313) 764-1555 or visit their Web site at http://www.aapor.org.

■ **The National Council on Public Polls** publishes another useful pamphlet, **Twenty Questions a Journalist Should Ask About Poll Results.** To obtain a copy, call (800) 239-0909.

■ **The Research Industry Coalition, Inc.,** publishes a brochure, **Integrity and Good Practice in Marketing and Opinion Research.** To obtain a copy, call (516) 928-6803.

■ **The Council of American Survey Research Organizations** publishes a pamphlet, **Surveys and You.** To obtain a copy, call (516) 928-6954, or visit their Web site at http://www.casro.org.