

Common Pitfalls in Conducting a Survey

Table of Contents

Time and Cost Requirements	1
Initial Planning	2
Questionnaire Design	3
Sampling Procedures	4
Pretesting	5
Nonrespondents	6
Processing the Data	7
Glossary	7

The intent of this brochure is to provide a brief review of potential problems that may occur during the survey research process. All of these common pitfalls can bias or invalidate the survey results. For a more thorough discussion of these pitfalls, please consult the list of technical references shown in the box below.

Time and Cost Requirements

One of the more common pitfalls is to underestimate the amount of time required to properly conduct a survey. There are many different stages involved in conducting a survey—identifying the information needs; defining and locating respondents; deciding how the data will be collected; designing and pretesting the questionnaire; hiring and training staff or contractors; determining how to handle non-response; and coding, cleaning, tabulating, and analyzing the data. Although some of these steps can be performed simultaneously, the entire process can be

quite time consuming. By not allowing adequate time, it might be tempting to skip one or more of the stages and inadvertently introduce errors or bias into the results.

The amount of time needed to conduct a survey depends on the complexity and type of survey. According to the American Statistical Association, it may be possible to conduct a simple, brief telephone survey in two or three weeks, but most surveys will take considerably more time. In their pamphlet, *How to Plan a Survey*, they state that several months to a year is a more typical timeline for most survey projects.

As a general rule, surveys are an expensive method of obtaining information. Sponsors of surveys should be wary of firms offering to conduct surveys at prices markedly below those of other firms. Often these “deals” result in second-rate work because the firm skips important steps in the process, violates proper sampling procedures, and does not employ adequate quality control



Technical References:

American Statistical Association
1429 Duke Street
Alexandria, VA 22314
703-684-1221
<http://www.amstat.org>

American Association for Public
Opinion Research
P.O. Box 14263
Lenexa, KS 66285
913-310-0118
<http://www.aapor.org>

Council of American Survey
Research Organizations
3 Upper Devon
Port Jefferson, NY 11777
631-928-6954
<http://www.casro.org>

Federal Committee on
Statistical Methodology
Statistical Policy Working Papers
<http://www.fscsm.gov>

procedures. The cost of a survey generally increases with the length and complexity of the questionnaire, the size and complexity of the sample, and the amount and complexity of the data analysis.

Initial Planning

Not clearly defining the purpose of a survey is a common pitfall. Before beginning a survey, it is strongly recommended that the following questions be answered:

1. *What is the purpose of the survey?*

The goal could be to seek specific information, predict future behavior, or test a particular hypothesis. Whatever the purpose, it needs to be clearly and concisely stated.

2. *Can the information be obtained by other means?*

It is quite possible that another organization has the desired information. Perhaps, published documents from the Federal Government or private institutions may suffice.

3. *Can the intended audience be reached?*

If there is no information that allows the researcher to define and select the members of the population to be surveyed, then it may not be possible to conduct a statistically reliable study that can be generalized to the entire population. For example, there may be no information on where persons with a specific characteristic live. Without location information, a simple random sample can be conducted in the hope of reaching enough of

the desired population. However, if the desired population is a small proportion of the general population, a very large sample would be needed and this may not be financially feasible.

4. *If a survey is needed, who will use the information and how will it be used?*

If the survey results will be made available to the public, if subpopulations are to be analyzed separately, or if a crucial decision will be based on the information, the issue of an acceptable

margin of error or precision needs to be addressed. Once the required accuracy level is decided, it will have a direct bearing on the sample size and the proper design of the questionnaire.

5. *Are specific cross-tabulations on certain characteristics desired?*

If so, the questionnaire should be designed in such a way that the desired information can be extracted easily from the results. The analyst doing the computer programming should be consulted about whether the desired cross-tabulations can be performed based on the questionnaire's design.

6. *Given the resources available, is the study feasible?*

The scope of the project may have to be diminished or the project postponed if money, manpower, and time constraints are binding.



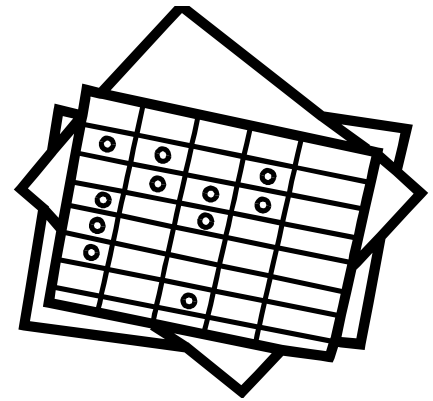
Questionnaire Design

The questionnaire should be designed with the study objectives in mind. The challenge for the researcher is to design the questions and word them in such a way as to obtain unbiased information. This sounds much easier than it is in practice. Careful attention must be paid to avoid sentences with multiple meanings; ones containing technical jargon; and lengthy, difficult, or sensitive questions. It is not uncommon for the sponsor conducting the survey to inadvertently introduce bias in the question design through the wording or ordering of the questions. For example, a series of questions querying the respondents on their concerns about crime followed by a question asking if more tax dollars should be spent on public safety is likely to result in a biased assessment. To counter this tendency, the survey should be reviewed by a diverse group.

Another common pitfall involves the use of open-ended questions rather than multiple choice. With open-ended questions, the range of answers

may be so varied that any comparisons will be meaningless. Open-ended questions cannot be handled easily by computer processing methods, especially with a large number of questions or a large sample size. If open-ended questions are used with personal interviews, the interviewer may not code the verbatim response and the failure to do so could change the entire meaning of the answer. It is strongly suggested that open-ended questions be used only when the question cannot be written in a multiple choice format. The obvious advantages of using multiple choice questions are the ease in answering, coding, and analyzing the responses. However, care must be taken to see that all the possible response alternatives are included and that these alternatives do not overlap.

When developing questions for a survey, abbreviations and colloquialisms should be avoided. Abbreviations and colloquialisms often are misunderstood or interpreted differently depending on one's



experiences. A question asking how frequently family members drink soda may be interpreted differently depending on the background of the respondent. One respondent may interpret this question as asking about soft drinks while another may interpret the question as asking about ice cream drinks.

Asking too much information in a single question can result in problems. Questions that require a respondent to make calculations such as the percent of income spent on housing costs are likely to elicit incorrect answers from respondents who do not know how to calculate a percent. It is preferable to ask two questions—one for annual income and a second for monthly housing costs. The analyst tabulating the data can

then calculate the desired percent from these two questions.

Requiring respondents to answer questions on items for which they have incomplete or little knowledge is a common pitfall. People often answer questions even if they do not have the knowledge the researcher requires. This happens because respondents define the question differently than the researcher, because respondents do not wish to appear ignorant or because respondents do not want to disappoint the researcher. A question asking a respondent if they have been to a Fairfax County park during the past year may elicit a "yes" answer even if the respondent has not

been to a county park. This may happen because the respondent may not know which parks are county parks and which are private, federal, state, or regional parks but they do know that they recently visited a park.

In mail or written surveys, interviewer bias is not a concern. When conducting a personal interview survey or telephone survey, the interviewers must be carefully trained so as to avoid possible bias and avoid influencing the responses. For example, if it is crucial to interview a particular respondent, like the head of the household, the interviewer must be informed of the possible bias of using a substitute respondent.

Sampling Procedures

It is rarely feasible to survey the entire population and, thus, a sample of the population is selected to be surveyed. Probability samples are the only type of samples where the results of the survey can be generalized to the entire population. In a probability sample, all the members of the target population have a known, non-zero chance of being included in the sample. Only probability samples allow the researcher to calculate the precision of the estimates and to specify sampling error. These statistics should be

made available to users of the data so they can evaluate the usefulness of the results. Because nonprobability samples are easier to conduct, it is often tempting to use a nonprobability sample when a probability sample is what is really required. Examples of nonprobability convenience samples are stopping customers in a store or surveying visitors to a web site. With these methods, valid generalizations cannot be made beyond the people interviewed. Also in most



cases, judgment sampling, where interviewers or "experts" choose those believed to be the most representative of the population, should be avoided. Different researchers are very likely to have different ideas about who the typical subject should be.

A probability sample, however, is only as good as the population from which it is drawn. It is a common pitfall not to spend the time necessary to ensure that the population from which the sample is drawn is complete and well defined. Often members that should be part of the population are excluded because of the difficulty of finding or reaching those members. For example, a telephone survey cannot reach all households because some households do not have telephones. Therefore, a telephone survey can only be generalized to those households who have telephones unless other means are used to reach households without telephones.

In addition, the researcher needs to be aware of other characteristics of a population

Pretesting

Following the questionnaire design and sample selection, a pretest should be conducted on a subset of the target population to find out if the questions elicit the desired

information. Although pretesting is frequently overlooked, it is the only way to discover possible misinterpretations and misunderstandings which may



Using a telephone survey as an example once again, most telephone samples are drawn using a list of telephone numbers or through random digit dialing techniques. Households that have more than one telephone line have a greater probability of inclusion in the sample than those households with only one telephone line. Thus, the researcher must adjust for these different probabilities of inclusion when sampling or else when analyzing the data.

information. Although pretesting is frequently overlooked, it is the only way to discover possible misinterpretations and misunderstandings which may

Often these adjustments are not done and this introduces bias in the data and conclusions.

Another commonplace pitfall is the substitution of available respondents when the selected respondents cannot be easily reached. Unfortunately when this is done, the sample ceases to be a true random probability sample. Substitution builds sample size without diminishing nonresponse bias.

A similar sampling pitfall involves enlarging the sample size to get “enough returns” despite an anticipated low response rate. A sample of 10,000 that elicits 1,000 responses is less reliable than a sample of 200 that elicits 190 responses. With low response rates, the researcher cannot be sure that those who do not respond are not very different from those who do respond.

arise from the questions. In addition, pretesting is the only way to discover problems that may arise from the field procedures to be used. Ideally, the pretest should test

all phases of the survey methodology—the question wording, the procedures to be used to conduct the survey, and the procedures to be used to analyze the results. Even the most experienced researcher cannot foresee all the potential problems that may occur when using new questions or

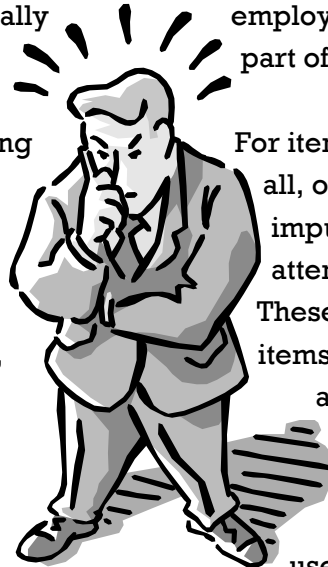
techniques. With a mailed pretest, be especially alert to a large number of "don't know" responses, which might signal a question which is too lengthy or difficult. Pretesting can be done a variety of ways. Some of these techniques include using focus groups, laboratory interviews, respondent

debriefings, interviewer debriefings, split panel tests, and analyzing item nonresponse. Because each of these pretesting techniques has strengths and weaknesses, it is not unusual to combine several of these methods when conducting a pretest.

Nonrespondents

Another commonly overlooked area concerns the issue of nonrespondents. According to the American Statistical Association, a low response rate produces more questionable results than a small sample, since there is no scientifically valid way to infer the characteristics of the population that the nonrespondents represent. So, rather than simply ignoring the nonrespondents, every attempt should be made to get them to respond to the survey. If further correspondence or callbacks do not substantially increase the response rate, weighting adjustments can be used to help compensate for potential nonresponse bias. For instance, if the response rate is lower among low income groups, weighting up the low income respondents to conform to a known income distribution might help eliminate nonresponse bias. However, if there are differences between the respondents and nonrespondents in each income group, then nonresponse bias will remain or increase despite attempts to reduce it through data weighting. Therefore, weighting adjustments

should not be considered a complete solution to the problem. A diligent effort should be made to obtain a high response rate to the survey. High response rates are usually obtained by employing rigorous follow-up procedures as part of the survey methodology.

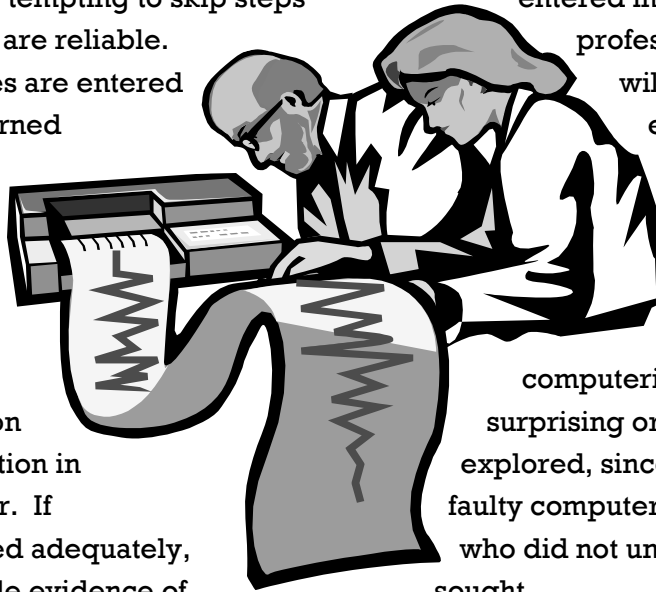


For item nonresponse, where some, but not all, of the questions are answered, various imputation methods can be used in an attempt to compensate for possible bias. These methods use the responses to other items in the questionnaire as information in assigning values to the nonresponse items. A major benefit of imputation is that a data set is constructed with fewer or no missing values. However, users of the data should be alerted that the data set contains imputed values, so that greater precision will not be attributed to the survey estimates than is warranted. The techniques discussed in this section can be useful, but they can be extremely difficult to implement. Therefore, only the experienced surveyor should attempt to employ these methods.

Processing the Data

The process of data entering, editing, and analyzing the completed questionnaires is subject to a great deal of human error. Maintaining a high level of quality control during these processes takes time, and it is very tempting to skip steps that ensure that the data are reliable.

Before the questionnaires are entered into a database, the returned questionnaires should be scanned to determine if respondents are leaving comments that could indicate that they misunderstood a question or are answering a question in an inappropriate manner. If pretesting was performed adequately, there should be very little evidence of



confusion on the questionnaires.

Ideally, all of the questionnaires should be checked to see if the responses match what was entered in the computer. Many

professional data entry organizations

will double enter the data to ensure accuracy. In addition,

further editing must be performed to check for impossible or inconsistent answers. This editing usually is accomplished by using

computerized crosschecks. Also,

surprising or unintuitive results should be explored, since these could be caused by faulty computer programming or respondents who did not understand what was being sought.

Glossary

- *Bias (error)*: Distorted or unreliable survey results. All surveys contain some bias. Bias is increased when the respondents (persons answering the survey) are not representative of the population being questioned, when questions are poorly written or misunderstood, and when the researcher uses inappropriate techniques to analyze the data.
- *Census*: A study using all available elements or members of a population.
- *Data*: The collection of observations and information resulting from the survey process.
- *Element (member)*: The basic unit about which survey information is sought (i.e., person, business, household, car, dog, et cetera).
- *Heterogeneous*: A population whose elements have dissimilar characteristics. Heterogeneity is the state of being dissimilar.
- *Homogeneous*: A population whose elements have similar characteristics. Homogeneity is the state of being similar.

- **Instrument:** The tool or device used for survey measurement; usually a questionnaire.
- **Nonresponse:** Unit nonresponse refers to the refusal of persons selected to be sampled to participate in a survey (i.e., person does not return the mail questionnaire). Item nonresponse refers to selected questions left unanswered by the person responding.
- **Population:** The universe or collection of all elements (persons, businesses, et cetera) being described or measured by a sample.
- **Pretest:** An initial evaluation of the survey design by using a small, subsample of the intended population for preliminary information.
- **Questionnaire:** A measuring device used to query a population/sample in order to obtain information for analysis.
- **Respondent:** An element or member of the population selected to be sampled.
- **Sample:** Any portion of the population, less than the total.
- **Sampling Frame:** An exhaustive list of all members of the population from which a sample can be drawn.
- **Survey:** A process of inquiry for the purpose of data collection and analysis using observation, polls, questionnaires, and/or interviews.
- **Statistics:** Descriptive measures based upon a probability sample.

Fairfax County Department of Systems Management for Human Services

Economic and Demographic Research
12011 Government Center Parkway
Suite 222
Fairfax, Virginia 22035

Phone: 703-324-4519
TTY: 711 (Virginia Relay)
Fax: 703-803-8598

Building Upon the Strengths of a Diverse Community



We're on the Web!
www.fairfaxcounty.gov/aboutfairfax