

ARF CRITERIA for Marketing and Advertising Research

Advertising Research Foundation



TABLE OF CONTENTS

	Page
INTRODUCTION	7
FUNCTION 1: THE ORIGIN AND PURPOSE OF THE STUDY.....	9
A. Who is the sponsor of the research?	
B. Who is the primary research firm?	
C. What is the purpose of the study?	
D. What is the background for the study?	
E. If action standards or criteria for decision making are included, are they stated clearly and explicitly?	
FUNCTION 2: THE RESEARCH PLAN	11
A. Was the overall concept and design of the study sufficient to accomplish the objectives?	
B. Was the context of the study adequately realistic?	
C. Was the timing of the study appropriate for the objectives?	
FUNCTION 3: THE SAMPLE.....	13
A. What type of sample design was used—probability or non-probability?	
B. What was the defined universe (population) studied?	
C. What type of sampling frame was used?	
D. What procedures were specified for selecting sampling units from the sampling frame?	
E. Was the sampling plan fully executed? Were provisions made for non-response?	
F. What was the sample size?	

Copyright © 1984 by Advertising Research Foundation, Inc.

Advertising Research Foundation
3 East 54th Street
New York, N.Y. 10022

June 1984

FUNCTION 4: DATA COLLECTION: THE QUESTIONNAIRE 16

- A. Was the questionnaire too long?
- B. Did the sequence of questions have a biasing effect?
- C. Was the questionnaire structured and formatted for efficient and accurate administration?
- D. Were the instructions clear and accurate?
- E. How was the questionnaire introduced or positioned to the respondent?
- F. Were the questions phrased clearly and objectively?
- G. Were the questions appropriate to the respondents being surveyed?
- H. Were repetitive portions of the questionnaire written in a consistent fashion?
- I. Was the questionnaire pilot tested?
- J. Were the questions appropriate to the intended analytical methods?

FUNCTION 5: DATA COLLECTION: THE FIELDWORK..... 21

- A. What training did the field staff receive for the study?
- B. Were quality control procedures employed to check the fieldwork in process during the study?
- C. Was the fieldwork verified?
- D. Was every effort taken to insure an acceptable response rate?

FUNCTION 6: DATA PROCESSING..... 23

- A. Was the questionnaire carefully edited?
- B. Was the coding system reliable?
- C. Was data entry verified?
- D. Were the tabulations checked?

FUNCTION 7: TREATMENT OF DATA..... 25

- A. How, if at all, were the in-tab sample data weighted?
- B. Were the effects of weighting indicated?
- C. How, if at all, were the data otherwise adjusted?
- D. Were assumptions for the use of statistical techniques justified?

FUNCTION 8: PRESENTATION OF SURVEY FINDINGS 28

- A. Are the results fairly presented? Do the data support the interpretations and conclusions? Are the actual findings clearly differentiated from the interpretation of the findings?
- B. How is the response rate presented?
- C. Are the complete findings presented?
- D. Is there a distinction made between association and causation?

FUNCTION 9: SUMMARY OF LIMITATIONS..... 30

- A. Is the study reliable?
- B. Is the study valid?
- C. Are the measures and experimental design sufficiently sensitive?
- D. Can the results be generalized?

ARF RESEARCH PRACTICES COUNCIL

Babette Jackson, Chair J. Walter Thompson Company

Contributing Members:

Joel Axelrod	BRX, Inc.
Lionel C. Barrow, Jr.	Howard University
Mark Clements	Mark Clements, Inc.
Ronald Gatty	Baruch College
Charles Jacobson	Life Savers, Inc.
Charles Lehman	Newspaper Advertising Bureau
Howard Leonard	Young & Rubicam, Inc.
Marshall Ottenfeld	D'Arcy-MacManus & Masius
Guy Parker	Cunard Line
Gilbert Sabater	Becket

The Council wishes to acknowledge the assistance of the ARF Board and Staff in final preparation of the *Criteria* and of Patricia Murray Finn for editorial review.

INTRODUCTION

This document is an update and extension of an earlier ARF publication also entitled *Criteria for Marketing and Advertising Research*, originally published in 1953 and reprinted many times since. These criteria are intended to provide helpful guidelines for planning and reporting market research studies and for assisting practitioners and users of market research to evaluate the quality, relevance and utility of the information reported.

Although this document is more extensive than the earlier one, no set of criteria can ever be complete or cover all situations and contingencies, present and future. Still, no matter what the situation, the underlying quality of the research requires careful attention to the basic functions outlined here.

The criteria for research presented in this report are discussed under the headings of nine functions. Function 1 is concerned with questions related to the origin and purpose of the study. Function 2 discusses the research plan. Functions 3 through 8 are concerned with the specific design and methods used to conduct the study. Function 9 covers questions concerning the overall quality of the study.

This new criteria report was developed by the members of ARF's Research Practices Council, under the chairmanship of Babette Jackson of J. Walter Thompson. The experience of the members of this Council represents all areas of applied marketing research, and this report was designed to draw on and summarize this experience for use by all those concerned with supplying and applying marketing research study information.

FUNCTION 1

THE ORIGIN AND PURPOSE OF THE STUDY

Each research report should present information about the identity of and background for the study, answering the following questions:

- A. Who is the sponsor of the research?
- B. Who is the primary research firm?
- C. What is the purpose of the study?
- D. What is the background for the study?
- E. If action standards or criteria for decision making are included, are they stated clearly and explicitly?

A. *Who is the sponsor of the research?*

In published research reports, readers have the right to expect the sponsor's identity to be disclosed in the report. In large-scale studies with multiple sponsorship, where it would be impractical to identify all sponsors either on the report cover, or even in the report, a list of the sponsors should be readily available to all users of the report on request from the company publishing the results.

B. *Who is the primary research firm?*

Identification of the primary research firm responsible for the design and the conduct of the study should also be clearly stated in the report. In using a research report, it is important to know the extent of the research firm's role and responsibilities in conducting the study.

C. *What is the purpose of the study?*

The first section of the research report should contain a statement of purpose which describes clearly what the study was meant to accomplish. The purpose of the study need not dictate one particular research design, data collection instrument or analytical treatment, but all aspects of the study should be cohesive and appropriate to the purpose.

D. *What is the background for the study?*

The report should also provide a general introduction to the situations, occurrences or conditions which led to the decision to do the research.

- E. *If action standards or criteria for decision making are included, are they stated clearly and explicitly?*

If the research is to provide the basis for explicit actions to be taken or decisions to be made, it is also important to know whether the action standards or criteria for decision making have been stated explicitly. This qualification applies most often to proprietary research undertaken for a specific sponsor rather than to published research reports meant for a broad range of application by users of the research. The stringency of the criteria should be appropriate to the purpose of the research and the business risk involved in the decision to be taken.

FUNCTION 2

THE RESEARCH PLAN

A research report should provide a general understanding of the overall research plan used in the study, including identification of the universe and sampling frame, the sample design features, the mode of data collection, the data collection methods, and the timing and geographic dimensions of the study.

Facts about the study's plan should be found in the report's introductory sections. If the study plan facts are treated in summarized form at the start of the report, fuller treatment should be available in a later section or, when necessary, in a separate technical document.

Issues which relate to the value of the overall research plan include the following:

- A. **Was the overall concept and design of the study sufficient to accomplish the objectives?**
- B. **Was the context of the study adequately realistic?**
- C. **Was the timing of the study appropriate for the objectives?**

- A. *Was the overall concept and design of the study sufficient to accomplish the objectives?*

Research design is an art as much as it is a science, and several alternative approaches may satisfy the objectives of the study. To assess the research design is to compare alternative approaches in order to judge if the approach selected is sufficient, given the constraints of time, budget, the potential value of the results and the potential risk of incorrect information. It is important to recognize that different measurements yield different types of data and are amenable to different kinds of analysis. Study design should insure that the data and analysis implied by the approach selected are appropriate to the research objectives.

- B. *Was the context of the study adequately realistic?*

Advertising and marketing researchers may adopt a questioning procedure and use a person's memory to assess past behavior, or they may choose to create a test situation in which behavior can be measured. If a test situation is to be created, behavior is ideally measured under real life conditions where people react, respond and make decisions without consciousness of being measured. For the most part, this is not feasible primarily because "real life" conditions inherently involve many unpredictable and, therefore, uncontrollable variables which make it difficult, and often impossible, to measure the particular behavior which is the subject

of research. This circumstance makes the practice of marketing research and, specifically, the use of experimental design, necessary.

In examining research involving a departure from "real life" conditions, consideration should be given to the degree to which the measurement is based on an analog of actual circumstances: Did the conditions adequately reflect "real life", or offer choices to respondents that correspond to the kinds of decisions that they might normally make? In other words, to what extent do the conditions of the study relate to the actual situation of the market? Does the trade-off between experimental controls for reliability of measurement and the departure from reality limit the actionability of the measurement? In experimental designs, are test conditions evident to the respondent, thereby introducing the possibility of response bias?

C. *Was the timing of the study appropriate for the objectives?*

The period of time when the study was conducted can have a profound effect on the value or meaningfulness of the data. For example, biases may arise because of seasonality, day of week or time of day in which data were gathered. In addition, the accuracy of responses in a study will be affected by the time interval between the study and the event being measured.

FUNCTION 3

THE SAMPLE

Information about sample design, size, and implementation should be presented in each research report. In evaluating the sample for a study, the first step is to examine the sample design—that is, how the sample was intended to be drawn. The second step is to examine how well the intended sample plan was executed.

Key questions concerning sample design and execution are as follows:

- A. What type of sample design was used—probability or non-probability?
- B. What was the defined universe (population) studied?
- C. What type of sampling frame was used?
- D. What procedures were specified for selecting sampling units from the sampling frame?
- E. Was the sampling plan fully executed? Were provisions made for non-response?
- F. What was the sample size?

A. *What type of sample design was used—probability or non-probability?*

There are two basic types of sample designs—probability samples and non-probability samples. There are many variations of each basic type, but the essential difference between the two types is that in a probability sample, the sampling units at every level (e.g. persons, households, families, purchases, stores, business firms, etc.) are selected on a random basis, while in a non-probability sample, the sampling units at one or more levels are selected in a non-random way.

It is desirable and conventional to use a probability sample for any survey which is designed to estimate the absolute incidence and dimensions of a characteristic or behavior within a defined population—for example, to estimate the number of adults who smoke cigarettes in the U.S., the number of cigarettes they smoke and the brands they smoke; or to estimate the proportion of persons who watch television on the average evening, the programs they watch and the amount of time they spend watching.

Non-probability samples are sometimes used when the goal of the study is something other than an absolute estimate of a characteristic or behavior in the population. For example, non-probability samples are often used for small-scale studies designed to explore reactions people have to a new product idea or brand perceptions that people obtain from a new television commercial. Non-probability samples may also be used in large-scale studies designed to track characteristics

and behaviors over time in a defined population, where the goal is not so much to measure the absolute incidence at any one point in time but rather the direction of change up or down in the characteristic or behavior from time to time—for example, specific changes in brand usage or consumer behavior. It is understood that measures of these changes from a non-probability sample are not projectable to the population.

In concept, probability samples are always more acceptable than non-probability samples. Probability samples are the only type of samples for which standard errors, or measures of precision, can be calculated. Although tests of significance are often applied to data from non-probability samples, such tests are not appropriate.

B. *What was the defined universe (population) studied?*

The universe for a study is the specified population to be represented. The universe for marketing and advertising studies is often defined in terms of people—for example, adults 18 years of age and over living in the U.S.—but it can also be defined in terms of other types of entities, such as households, families, occurrences of certain behaviors (e.g. purchases), retail stores, business firms, etc. The exact criteria for defining the survey universe may sometimes seem trivial, but they can vitally affect the resultant data; different criteria used in different studies may make a comparison of results impossible.

C. *What type of sampling frame was used?*

The sampling frame for a study provides the means of access to the defined universe, enabling the identification of sampling units. The sampling units for a study are the specific people, families, households, retail stores, business firms or other survey entities selected from the universe to represent the universe for the study.

The sampling frame may be a list or set of files, such as telephone directories, files of charge-account customers or subscription lists. Or the sampling frame may be a set of maps along with population counts for units that can be located geographically; or traffic flow records, such as transactions at a supermarket check-out.

Since the sampling frame is the base from which sampling units are drawn, for probability sampling it is important that the frame be appropriate for the defined universe and as complete and current as possible. Sampling frames are generally imperfect, requiring that procedures be built into their use to correct for the imperfections. For example, maps may be out of date and not show new housing developments; non-telephone and unlisted telephone households do not appear on telephone directory listings; and some names and addresses on subscriber lists may be inapplicable or incorrect.

In the use of telephone samples, it is also now conventional to correct the sampling procedures by supplementing the sample based on telephone directory listings with a sample of unlisted households obtained from a random digit selection process, or by using telephone samples derived completely by random selection. This process obviously does not correct for the exclusion of non-telephone households for a universe defined as all households.

D. *What procedures were specified for selecting sampling units from the sampling frame?*

There are many procedures which can be applied in selecting sampling units. Each procedure has specific cost and precision implications that should be taken into account in planning research and in using research reports. In this area, it is also important to recognize the relationship between the mode of interviewing and sampling selection. The mode of interviewing has to be appropriate to the accessibility of the sample universe.

E. *Was the sampling plan fully executed? Were provisions made for non-response?*

The sampling plan may call for a distribution of interviews across time, across specific geographic definitions, across sampling clusters, etc. In the execution of the sampling design, care must be taken in the planning and fieldwork to insure that the sampling plan can accomplish the distribution specified and then, the execution of the distribution should be checked, from the fieldwork, to examine the in-tab sample for: control of sample selection by the central office of the research company; distribution of the in-tab sample across time, geography, or by other sample specifications, during the survey period; and size and completion of sampling clusters.

An important summary measure reflecting the adequacy with which the sampling plan is carried out is the response rate for the study. Response rate is the proportion of sampling units originally designated for the sample that actually provide information for the research. Because of the potential bias introduced by non-response, it is desirable for the study design to include provisions for overcoming non-response. Some of the methods used to overcome non-response in a survey include call-backs or repeat mailings, incentives, and designating a sub-sample of non-respondents for concerted follow-up.

F. *What was the sample size?*

The adequacy of sample size in any study is associated with the type of behavior being measured, the information goals for the study (for example, how many sub-groups are to be isolated and examined), and the estimated precision required in the study's results. Decisions about sample size in a study plan may be affected by constraints of time and cost.

The adequacy of the sample size for probability samples can be gauged by standard error statistics for estimates derived from the study. For many probability sample designs, the standard error of the estimate is inversely proportionate to the square root of the sample size. The smaller the standard error the greater the precision. However, not all sample designs are equally efficient. Other features of the design, such as stratification or clustering, can increase or decrease the degree of precision.

In some studies there is disproportionate sampling or supplementation of the sample in order to provide a sufficient number of cases representing a particular sub-group. In such cases, the size of the general sample without over-sampling, the size of the sub-group sample and the equivalent simple random sample size should be reported.

FUNCTION 4

DATA COLLECTION: THE QUESTIONNAIRE

The adequacy of data collection in any study is dependent on the quality of the data collection instruments used and on the quality of the fieldwork in which the instruments are used to obtain information from respondents (sampling units). Considerations concerning fieldwork are discussed under Function 5.

There are various types of data collection instruments including: questionnaires administered by interviewers, questionnaires which are self-administered by respondents, diary instruments used in media audience research and in product use or other behavioral studies, and discussion outlines used in focus groups or other qualitative work. Other data collection instruments include various mechanical devices such as meters used to measure television set usage, eye movement cameras used to measure response to packaging and advertising stimuli, and UPC supermarket check-out records used to monitor product purchases, etc.

In addition to the main data collection instruments used in a study, other supplementary tools or materials may be needed for the questioning process. These might include various stimuli needed during the interview such as new product samples, television commercials, print advertisements, television program pilots, prototypes of new products, magazines or newspapers, printed concept statements (usually presented to the respondent on separate show cards), logo cards used in magazine audience research, and stripped issues of magazines used in magazine research. Supplementary materials also include various types of response cards and other response devices such as sort boards, or response dials and buttons.

The user of any research report should obtain specific information about the data collection instruments used for the study and should examine the instruments and materials when feasible. That is, the report user should have a clear understanding of how information was obtained from the respondents and should have a clear perception of the respondent's task involved in the study. A questionnaire, a copy or a description of other data collection instruments should be included in all research reports, or if this is not feasible, they should be readily available from the research company or sponsor. Supplementary materials are usually not presented in the study report, but they should be readily available to the report user on request.

In evaluating a questionnaire (or other data collection instrument) it is important to address issues: (1) concerning the structure and thrust of the questionnaire as a whole; and (2) concerning the content and structure of the individual questions which make up the questionnaire.

Key issues concerning the questionnaire include the following:

- A. Was the questionnaire too long?
- B. Did the sequence of questions have a biasing effect?
- C. Was the questionnaire structured and formatted for efficient and accurate administration?
- D. Were the instructions clear and accurate?
- E. How was the questionnaire introduced or positioned to the respondent?
- F. Were the questions phrased clearly and objectively?
- G. Were the questions appropriate to the respondents being surveyed?
- H. Were repetitive portions of the questionnaire written in a consistent fashion?
- I. Was the questionnaire pilot tested?
- J. Were the questions appropriate to the intended analytical methods?

A. *Was the questionnaire too long?*

An overly long questionnaire can lead to a respondent's answers becoming increasingly careless and superficial, to the point that the data collected are often inaccurate or incomplete. Excessive questionnaire length can also negatively affect the way in which an interviewer will conduct an interview.

In considering questionnaire length, it should be recognized that the dimension of questionnaire length has both an objective and subjective component. A questionnaire with many pages taking an hour to complete may subjectively be considered short if it is interesting, involving and stimulating; conversely, an uninteresting, uninvolved questionnaire that consumes 15 minutes may seem interminable. Locale and circumstances may also affect the perceived length of a respondent's task.

Decisions about questionnaire length are made by the researcher based on the information needs of the survey, based on the judgment and experience of the researcher in conducting similar studies, and based on pre-trial of the questionnaire in pilot studies. Symptoms of excessive length can be reduced response rate overall or reduced completeness of in-tab questionnaires.

B. *Did the sequence of questions have a biasing effect?*

The order in which questions are asked can have a significant influence on the results. The sequence should establish a natural, logical flow for the respondent.

The positioning of one question versus another may condition responses. For example, a brand recall question asked first may affect a brand preference question asked subsequently, and vice versa. It is important to know what the sequence was in such a case, and whether any rotation of questions or other device was used to lessen or avoid sequence effects.

The context in which a question is asked can also influence the results. For example, a product attribute may be given a higher rating if it is evaluated in the midst of a large number of poorly-regarded attributes; conversely, it may be given a lower rating if it is being contrasted with a set of highly-regarded attributes. The rating has meaning within the context of the study, but not in an absolute sense.

With any question asking about a number of entities (for example, a brand rating question about a number of brands or a magazine readership question about a number of magazines), the order in which the brands or magazines are presented can also affect the findings. Sequence biases of this kind can be allocated evenly by techniques such as rotation.

C. *Was the questionnaire structured and formatted for efficient and accurate administration?*

The physical structure of a questionnaire should be designed so that it can be administered efficiently and accurately. Some questionnaires are quite complex in order to accommodate alternative contingencies, skip patterns, or screening and funnelling of respondents to different sections of the questionnaire. These complications may make sense from the perspective of the survey designer, but they are often difficult for the respondent or interviewer to cope with, leading to possible errors in data collection. Pre-trial of the instructions and questionnaire, before the full study, can help spot problems in administering the questionnaire.

Computer-programmed questionnaires may assist in simplifying questionnaire administration.

D. *Were the instructions clear and accurate?*

There are two types of instructions in a typical study: those contained in a document separate from the questionnaire and those which guide a respondent or interviewer through the instrument and are contained therein.

Detailed instructions in an accompanying document typically explain to interviewers the purposes, methods and procedures of the study as a whole, as well as clarify any portions of the questionnaire that are not self-evident. It is essential that these instructions are clear, particularly if an in-person briefing does not take place.

Instructions in the questionnaire itself specify probes, skip patterns which direct the interviewer or respondent to different sections of the questionnaire (usually contingent upon a respondent's specific answers to previous questions), or instructions to the respondent explaining how specific questions or types of questions should be answered. It is important that these instructions can be easily and accurately followed.

E. *How was the questionnaire introduced or positioned to the respondent?*

The information a respondent has about non-questionnaire elements of the study can influence results of the study. Knowledge of such elements as the identity of the sponsor, the purpose of the study, the presence of a tape recorder or a viewer behind a one-way mirror, the degree of anonymity of the respondent and the confidentiality of response can all affect the way persons respond to questions.

In general, it is important that the respondent not know the identity of the sponsor of the study. The overall goal is to conduct the research in an objective context.

F. *Were the questions phrased clearly and objectively?*

There are two key sub-questions under this heading. The first is—were the questions asked in such a way that all respondents are likely to understand them in the same way? It is important that questions be stated clearly and unambiguously. If not, responses may be off target with respect to the intended meaning of the question and also with respect to variations in meaning among the various respondents. A classic example of an ambiguous question in the media research area is "which of the following magazines do you read regularly?" "Regularly" needs definition of time frame to be interpreted the same way by each respondent.

A second question concerning the phraseology on a questionnaire is whether the wording used may tend to bias a respondent to answer in a particular way. It is clearly important that if leading questions are asked, the resultant data should be recognized as biased information.

G. *Were the questions appropriate to the respondents being surveyed?*

It is important to assess whether the terminology of the questionnaire is understandable to the respondent. Vocabulary differences exist between different regions of the country, socio-economic levels, cultures, ages and ethnic groups, as well as between specialists and non-specialists. A questionnaire on health issues directed to the physician may employ a different vocabulary than one addressed to the general population. Physicians know about angina, while patients know about chest pains. Manufacturers know the indices of turbidity; consumers know how cloudy the product looks.

H. *Were repetitive portions of the questionnaire written in a consistent fashion?*

Changes in question wording that appear to be minor can produce answers that are quite different. Therefore, if similar kinds of information (e.g. brand ratings) are collected at different points in the interview, it may be advisable (given that there are no explicit reasons for a difference in wording) that questions be asked in the same way and recorded on the same type of scale. When the same data are collected at different times (e.g. in a tracking study), the questioning procedures should be identical over time in order to compare the data.

I. *Was the questionnaire pilot tested?*

To evaluate the combined characteristics of the questionnaire, it is advisable that the instrument be pretested in advance of the main study to determine whether it is clear, understandable, not too long (not fatiguing), and feasible in terms of administration and respondent task, and also to check whether the overall goals of the study are being met by the planned measurement approach.

J. *Were the questions appropriate to the intended analytical methods?*

Audits, group sessions, depth interviews, open-ended questions, checklists, rankings and ratings each provide different kinds of data, and are amenable to different

types of analyses. In some cases, a proposed analytical approach mandates that questions be asked in a particular way; conversely, certain question construction precludes the possibility of certain kinds of analyses. For example, content analysis usually requires open-ended verbal responses; however, such open-ended responses cannot readily be treated by a factor analysis.

FUNCTION 5

DATA COLLECTION: THE FIELDWORK

Fieldwork is the study element often furthest removed from the supervision and control of the study sponsor and his primary research supplier, and furthest removed from examination by the report user. It is to be noted that frequently fieldwork is accomplished within a multi-strata configuration. The primary research firm may have an internal field department or may use an independent field service organization. In turn such organizations may subcontract parts or all of the fieldwork to other similar organizations.

Increasingly, fieldwork may entail computer-assisted interviewing procedures, especially with respect to telephone interviewing. The report user should be alert to the technology applications available and their potential value for research.

The report user should seek information to answer the following questions about the fieldwork:

- A. What training did the field staff receive for the study?
- B. Were quality control procedures employed to check the fieldwork in process during the study?
- C. Was the fieldwork verified?
- D. Was every effort taken to insure an acceptable response rate?

A. *What training did the field staff receive for the study?*

The amount and type of training to be given to the fieldworker will vary by type of study. Since different types of studies (e.g. qualitative, telephone, personal in-home) require special training and experience, one might ask whether the fieldworker is qualified for or accustomed to the type of interview being used in the study and, if inexperienced, whether specific training for conducting the study has been given.

Briefing is necessary to insure that each fieldworker knows the specific details and problems that might arise in the conduct of the elements of the project. The content of the briefing will depend upon the sample design, content and complexity of the questions, and the way the field instruments are designed and administered.

B. *Were quality control procedures employed to check the fieldwork in process during the study?*

Quality control is important as it provides assurance that the data are gathered carefully and by planned procedures. To an extent, interview verification, dis-

cussed in the following paragraph, will meet this need. However, in addition to any verification program, there remains the need to insure day-by-day quality of fieldwork, especially for large-scale studies over extended time periods. Continuing checks should be made to insure that data collection procedures are followed and to provide objective evidence on how well the work is done.

C. *Was the fieldwork verified?*

The purpose of fieldwork verification is to determine whether the interview was actually conducted with the designated respondent and whether the proper mode of questioning was used. Sometimes checks are also made of factual information obtained in the interview, particularly respondent and household classification data.

Verification may be implemented by a third-party firm independent of both the sponsor and the research supplier for the study.

The percentage of fieldwork verified varies from study to study. Many studies have a minimum verification requirement for an interviewer's work. Then if errors are found for any interviewer, 100 percent of that interviewer's work is verified. Simultaneous monitoring and checking of telephone interviewing work, if properly done, may greatly reduce the need for verification.

D. *Was every effort taken to insure an acceptable response rate?*

The sampling plan is but a means to an end. It is the response of the sample which matters and which must be pursued and collected to obtain as full a response rate as is possible. As stated in the original 1953 ARF *Criteria* document: "... It is desirable to get as close to a 100% return as possible."

In practice, of course, response rates fall below the objective of 100% return. It is important that the report user consider the implications of non-response, both conceptual and statistical. Efforts to attain an acceptable level of response in order to diminish these implications are encouraged.

FUNCTION 6

DATA PROCESSING

The function of data processing is to translate raw data from the study into usable, reportable form. Data processing includes editing, coding and tabulating. Editing is the process by which raw survey data are corrected for data flaws such as inconsistencies or incomplete entries. Coding is the process by which raw data (after editing) are identified by response or analysis categories in preparation for tabulation. Tabulation is the process by which the coded data are counted in each response or analysis category. The basic principle that should be used in all of these processing functions is that the rules for each should be as objective as possible, and established prior to execution of the process.

Computer-assisted interviewing procedures may have major impact on the steps involved in data processing. The report user needs to be alert to the nature of this impact, if the traditional steps are altered.

The report user should obtain a general understanding of the data processing methods used in the study. In particular, the report user should be assured that the answer to each of the following questions is "yes."

- A. Was the questionnaire carefully edited?
- B. Was the coding system reliable?
- C. Was data entry verified?
- D. Were the tabulations checked?

A. *Was the questionnaire carefully edited?*

In the process of editing, a review of respondent answers is made to reduce errors, inconsistencies and incomplete responses in the questionnaire. The goal in the editing process is to do this while at the same time preserving the meaning and integrity of the collected data. The report should provide a review of major editing rules used in the study: for example, how "no answers" are handled and how multiple answers to questions asking for one answer are handled.

B. *Was the coding system reliable?*

Whether done manually or by the computer, the object is to provide a scheme whereby responses are objectively translated into a given set of categories. Most questionnaires are precoded, thus eliminating variations due to individual coders. Sometimes codes are too cumbersome or cannot be established before fieldwork is

completed. To insure that the data are uniformly summarized, coding instructions must be explicit. Sample checks of each coder's work can pinpoint deviations and permit corrections as needed.

C. *Was data entry verified?*

The transfer of data from questionnaires into data processing form should be checked. Duplicate entry by different operators will serve to reduce data entry errors. A sample of each operator's work should be verified.

D. *Were the tabulations checked?*

Trial runs of the computer program should include double checks to insure that the data are consistent and the tabulations accurate. Hand tabulations or intermediate computer tabulations of a subsample are also helpful in checking the computer program.

FUNCTION 7

TREATMENT OF DATA

The term "treatment of data" designates any mathematical or statistical operation performed on the raw data other than simple counting or calculation of percentages. There are two main categories of such data treatment. One type consists of various weighting functions performed to project the in-tab sample data to produce population estimates. The other type of data treatment consists of various types of data manipulations which are used to convert survey data into indices of behavior or composite scores, for example, or to adjust the survey data by some known factor as is done in some calibration methods. Key questions in this area are:

- A. How, if at all, were the in-tab sample data weighted?
- B. Were the effects of weighting indicated?
- C. How, if at all, were the data otherwise adjusted?
- D. Were assumptions for the use of statistical techniques justified?

A. *How, if at all, were the in-tab sample data weighted?*

When population estimates are desired, the sample data may or may not be weighted.

In projecting the results of sample surveys, the data may be subjected to three types of weighting. First, weights are used to account for probabilities of selection. If only one adult is selected per household, for example, the sampling fraction in one-adult households is one and the corresponding weight is one, while in two-adult households the sampling fraction is one-half and the weight is two. Where sampling is done in several stages, the probability weight for each stage is the inverse of the sampling fraction and the total weight is the product of these weights.

The second type of weight in projecting sample data are weights to account for non-response. For a study with an 80 percent response rate, the weight would be 1.25 (100/80) while for a study with a 60 percent response rate, the weight would be 1.67 (100/60). Such weighting is frequently done by sampling area or cell, thus allocating the non-response bias on a geographic or demographic basis.

The third type of weight used in projecting sample data are weights used to bring sample projections into line with known population estimates. This sample balancing is frequently done by demographic cell, such as age and education. When the sample is well planned and executed, these adjustments are minor. The user can observe the effect of such adjustment by asking to see the sample composition before balancing was applied.

The use of computers simplifies data weighting and permits simultaneous weighting for non-response and data adjustment. To reduce variations in weights, computers can also impose upper and lower limits to the weights applied.

Weighting can limit the depth of the analysis inasmuch as weights increase sample variation. This is particularly true when cross-tabulations are done. Because some of the cells considered may have been subjected to large weights, the differential among weights in the various cells may be substantial. In that case, the tolerances associated with the weighted data are also high.

In general, the second and third types of weighting should be used with caution. A best estimate of any phenomenon is obtained by using a high quality sampling technique and a carefully designed data gathering procedure which is completed with the highest possible rate of response. Given that the sampling technique and data gathering instruments are of that caliber and the response rate is in the range of 80 to 90%, it is difficult to "improve" upon any estimates.

Perhaps most important, the concept of "improvement" implies an objective standard. The advocate of weighting bears the burden of proof that "improvement" has occurred.

One problem with trying to "improve" an estimate by sample balancing is that the available population measures and the sample measurement of the characteristic(s) on which the balancing is to be done may be of questionable accuracy.

B. *Were the effects of weighting indicated?*

Summary facts about the effects of weighting should be available to the report user. These include information about the range of weights used and the effects of weighting on tolerance estimates. The report should also show the unweighted data bases for each sub-group cited in the analysis and for the total sample.

C. *How, if at all, were the data otherwise adjusted?*

Often, in reporting a survey, responses to several items or questions are combined to produce a single measure for each respondent. When this is the case, it is important to know which items or questions were used and how they were weighted.

The construction of a composite score may also involve other considerations. For example, the response categories provided may be in the form of a "scale" ("Strongly Agree," "Agree Somewhat," etc.). Technically, this is called an "ordinal" scale; it has no zero point and the intervals between categories may not be equal. Nevertheless, the construction of a composite score from such questions often assumes that these discrete categories are points on a continuum to which specific values can be assigned ("Strongly Agree" = 5, "Agree Somewhat" = 4, etc.).

A number of marketing research concepts are difficult to measure directly because of budget limitations or the impossibility of seeing what is going on inside people's minds. Among these are "influence," "persuasibility" and a variety of other labels attached to individual behaviors, family relationships, etc. In most cases, the concept is of interest mainly in the context of actual behavior, but the "measurement" is usually based on verbal responses to questions. Words can never be equated with behavior. Knowing the questions and how these measures were treated (if combined or indexed, etc.) is important knowledge to the report user.

D. *Were assumptions for the use of statistical techniques justified?*

The data must meet certain conditions in order to justify the application of a given statistical technique and the conclusions derived, especially when the sample is small or when the technique is applied to small sub-samples.

Analytical techniques such as regression, factor analysis, and multidimensional scaling involve underlying technical assumptions. The user is advised to explore with a statistical specialist whether the assumptions are satisfied.

FUNCTION 8

PRESENTATION OF SURVEY FINDINGS

Key questions concerning the presentation of findings in a research report are as follows:

- A. **Are the results fairly presented? Do the data support the interpretations and conclusions? Are the actual findings clearly differentiated from the interpretation of the findings?**
- B. **How is the response rate presented?**
- C. **Are the complete findings presented?**
- D. **Is there a distinction made between association and causation?**

A. *Are the results fairly presented? Do the data support the interpretations and conclusions? Are the actual findings clearly differentiated from the interpretation of the findings?*

It is important that the reporting of study findings and the conclusions drawn from them be clearly distinguished. Reporting of the research data should be expressed in language which clearly represents the findings and from which all readers may obtain the same understanding of what is presented.

Interpretations or conclusions based on findings may be subject to differing points of view, and are acceptable as long as the findings logically support that point of view.

B. *How is the response rate presented?*

As previously defined, the response rate is the proportion of sampling units originally designated for the sample that actually provide information for the research. In other words, the response rate is the percent of the predesignated units that are in the tabulated sample. Since the base is predesignated units, substitutes cannot be counted in the computation of response rates.

The computation of response rates is sometimes difficult, particularly when a unit must be contacted before it can be determined that it is to be designated for the sample. For example, for a study of the teen-age population, it may not be possible to determine whether there are qualified respondents (teen-agers) in those households that refuse access or where no one is at home. When random digit dialing is used for a telephone study of the household population, it may be difficult to determine whether no answers and busy signals represent businesses, households or phantom numbers. To meet such difficulties, working assumptions

are sometimes made to adjust calculations on the basis of data compiled from contacted units.

The report of the response rate should describe the method along with the numbers entering the computation. For multi-stage and stratified samples, it may be of interest that response rates be reported by stage and stratum. It is desirable, when there are key sub-groups for analysis, that response rates be given for each key sub-group for which data are available to make the computation. Sometimes only partial information is obtained from a sampling unit. Report of response rate should show the number and percentages that supplied complete information; that provided sufficient information for most analyses; that were rejected because of non-verification, incompleteness or illegibility; and that were not reached at all or refused.

C. *Are the complete findings presented?*

Although it is not necessary for a research report to contain a written statement of every item of data gathered during the study, all relevant items which are specifically related to the study's objectives should be analyzed and shown. This is true even if one or more of the items is inconsistent with the overall conclusions drawn from the research. Findings not included in the report should be available to the report user on request.

D. *Is there a distinction made between association and causation?*

Association and causation mean different things and therefore should not be used interchangeably.

In most research studies conducted, association is the proper terminology when doing a statistical analysis. For example, when analyzing differences in demographic sub-group levels of brand usage, associations are being examined. There are occasions, however, when causality is a proper term to use or infer during interpretation of the results. In controlled experiments where assignments of treatments are made at random, such as in advertising weight tests, causality can be attributed in the interpretation of data.

If the research objectives require a conclusion about causality as distinct from association only, the study procedure must be designed accordingly.

FUNCTION 9

SUMMARY OF LIMITATIONS

The overall quality of a study is based on the combined effect of the quality of design and implementation in each specific area of survey conduct. The following questions are a good basis for reaching an overall understanding and appreciation of a study's findings:

- A. Is the study reliable?
- B. Is the study valid?
- C. Are the measures and experimental design sufficiently sensitive?
- D. Can the results be generalized?

A. *Is the study reliable?*

The question of reliability is concerned with the precision of the study, in terms of how the results of the study, based on a sample, differ from what would be found based on a study of the entire universe. That is, what is the sampling tolerance or error?

The reporting of data from sample surveys should carry understandable and correctly calculated information on the statistical reliability of the major findings or a statement that the reliability cannot be computed. If probability sampling is used and an adequate response rate obtained, the margin of sampling tolerance in survey results can be estimated by the application of statistical methods, based on the mathematical theory of probability. Calculation of sampling tolerance margins should take into account the nature of the sampling design (e.g. clustering, disproportionate selection, correlated observations, etc.) as well as the size of the sample.

Sampling tolerance margins should be stated without implying that the type of error they treat is the only one that may affect the findings. The discussion of data reliability should not obscure possible questions about the overall accuracy of the results (including non-sampling as well as sampling errors).

B. *Is the study valid?*

That is, does the research measure what it purports to measure?

A major source of potential error made in marketing research is to measure one event while claiming to measure another. For example, studies typically measure recall of magazine readership but say they are measuring magazine readership. As another example, many research reports are based on measures of attitudes which are either explicitly or implicitly assumed to be indicative of purchase behavior.

It is important to look for evidence that the measured event has a known rela-

tionship to the real-world event in which the reader is interested. If there are no data concerning validity, then it is necessary for the reader to make a judgment as to whether what is measured should be accepted as an adequate surrogate.

C. *Are the measures and experimental design sufficiently sensitive?*

It is important to consider the possibility of insensitivity when examining a study proposal or, after a study is done, when data show no change or no difference. The reader should recognize that it is quite possible to have a measure which is both reliable and valid, but which is inappropriate for the situation in which it is used.

One example is the use of a "brand of insurance purchased last" question for evaluating the effect of an insurance advertising campaign. It may well be a reliable measure with predictive validity. But responses to an "insurance brand purchased last" question would not be sensitive enough (or have enough relevance to the effect of the current campaign) to be suitable, the reason being the nature of insurance purchasing and the purchase cycle involved.

D. *Can the results be generalized?*

A study may be reliable, valid and sensitive, but still not be generalizable. For example, the results of a product test, conducted in a sample of test markets, may discriminate, the results may be replicable, and the results may accurately predict what other people who shop in the markets prefer. But preferences in the category may be very localized, and the results may not reflect preferences in other markets.

All studies are, in some degree, artificial, and limited by the particular procedures used. The user must give serious consideration to all particulars before generalizing beyond the specific set of observations presented by the study.

ABOUT THE ADVERTISING RESEARCH FOUNDATION

The Advertising Research Foundation, a professional association of member companies representing all segments of the advertising business community, is dedicated to advancing the discipline, practice, professional standards and validity of advertising research.

The ARF is taking the initiative in establishing itself as a "research quality presence" in the industry through the activities of its recently-formed Research Quality Council, through annual workshops on the subject of research quality, and through the publication of this document. This role is consistent with the by-laws of the organization, which include the following objectives:

- to analyze and evaluate existing research methods, and to define their proper applications and limits of usefulness;
- to recommend research standards;
- to encourage the highest level of honesty and professional responsibility in the practice of research.