INFORMATIONAL NEEDS FOR CURRENT DEMOGRAPHIC SURVEY DESIGN WITH DISCUSSION OF KEY REDESIGN RESEARCH PROJECTS

Lawrence Cahoon, R. Maurice Kniceley, Jr., and Gary M. Shapiro
U. S. Bureau of the Census

I. INTRODUCTION

Currently, personnel at the U.S. Bureau of the Census are engaged in major survey-design and methods research required to complete the post-1980-census redesign of the current demographic surveys. Although current effort is primarily directed to redesign, we also must frequently establish the designs for new surveys. Since data requirements for initial design and redesign are quite similar for a survey, we will be covering the broad range of both survey design and survey redesign in addressing the issues of informational needs and research efforts at the Bureau in this paper.

It is essential that the sponsor realize that it is the survey design statisticians' responsibility to design the most efficient sample survey possible, i.e., to design one which will obtain maximum information with the smallest cost and error. We may all set as a goal, the design of the ideal sample survey, but we know in advance it will be unobtainable and that compromises will become necessary. Because the sponsor, user, and design statistician share interest in, and must be concerned about final results, decisions on compromises become a mutual responsibility. Thus, informational needs as described herein extend beyond a simple statement of data requirements to include the criteria which the statistician is to use in making and/or incorporating compromises in a manner that will produce the most efficient design those constraints will permit. To this end, success depends heavily upon the extent to which user and sponsor have provided timely, essential information to the design statisticians.

This paper addresses the issues related to information the statisticians need, particulary from the sponsor, to design or redesign a given survey within the constraints. Because of the predominant importance of good survey objectives, we have limited discussion of informational needs to that facet of survey design and describe in general terms why the information is needed. We, then, describe how the information is specifically being used in key redesign research now underway.

II. DETERMINATION OF SURVEY OBJECTIVES

By far the most important informational need for survey design is the description of survey objectives. This is established by the data user and survey sponsors and, other than an initial warning that a survey is needed, should be the first information provided to the Bureau. Every aspect of survey design should be related directly to, and procedures used should result from, interpretation of those objectives. Once the objectives of the survey are completely established, it is then possible for the survey sponsor and the design statisticians to work together to determine additional information required for designing and implementing the survey.

A description of survey objectives should contain four overlapping areas of information: (1) the subject of the survey, (2) the basic purpose of the survey and major decisions which sponsors and users can be expected to make based on survey results, (3) the different types of analysis and uses to be made of the data, and (4) a listing of major statistics, their relative importance, the primary level (e.g., geography, race) at which the estimates are to be used, and required reliability.

In our experience, the first area, the subject of the survey, is easily described. The second area, the basic purpose and major decisions needed, is easily described when the purpose is clear-cut and there are definite decisions that are needed. The third area, types of analyses and data uses, is usally not difficult to establish in general terms. But it is always very difficult to list specific key statistics, their relative importance, the level at which the estimate is to be published, and the required reliability. As a result there is a tendency for sponsors to present such things as the subject, purpose, and tabulations in general terms and to want to discuss specific design details, such as sample size and rotation plans, before establishing complete objectives of the survey rather than afterwards as they should be. The net result is that the specific objectives are provided too late to be fully reflected in determining the design details and implementation of the survey. For example, we are still working with the sponsors to prepare a fully-specified set of objectives.

Why is it so important that design statisticians have such information and have it so early? The discussion of sample size estimation and sample design refinements throughout the paper will answer why the information is needed. The answer to "why so early?" should be evident from the following discussion of the logical sequence of events from the time the sponsor has first approached the Bureau with a request for design of a survey to completion of the survey design.

As a first step in the sequence, we ask the sponsor to prepare a set of specific survey objectives to guide us in the design of the survey. As soon as the sponsor has complied, we review those objectives and meet with the sponsors for clarification on questionable points and to negotiate any changes both parties may believe are necessary to the objectives.

With information from the "revised" set of objectives, we calculate a rough estimate of the sample size required to fulfill those objectives. By this time the sponsors are becoming concerned about what the survey will cost so we use the rough sample size estimates and generate preliminary estimates of what the survey will cost.

The Bureau can now begin actual design of the survey with first effort being directed to determining what design refinements, such as stratification and clustering, might be implemented. Subsequently several meetings are held with the sponsor to report status of the project, to seek clarifying information on objectives, and to

seek guidance as to which of potential alternative methods to implement.

Due to its importance, sample size estimation will be mentioned throughout the paper as an example of why the information is required. To help in understanding those discussions, we have presented below the formulae we might use to obtain the rough estimates of sample size mentioned previously. The formula have been presented in one place rather than throughout the paper so that differences in inputs required for the different estimates are more readily apparent. Thus the need for specific information to be specified in the objectives also is more apparent. The input terms to the formulae are defined generally leaving specific descriptions to later discussions in the text. The formulae are, for

estimates of a mean value $$n_{\bar{x}} = \frac{k^2\sigma^2}{E^2} \quad (1)$$

estimates of a proportion $$n_{p} = \frac{k^2PQ}{E^2} \quad (2)$$

estimates of a total value $$n_{x} = N^2\frac{k^2\sigma^2}{E^2} \quad (3)$$

estimates of a total number $$n_{a} = N^2k^2\frac{PQ}{E^2} \quad (4)$$

estimates of mean or total value when error is a percentage $$n_{\bar{x}(Rel)} \text{ or } n_{x(Rel)} = \frac{k^2V^2}{E^2} \quad (5)$$

estimates of a proportion or total number when error is a percentage $$n_{p(Rel)} = n_{a(Rel)} = k^2\frac{Q}{PE^2} \quad (6)$$

where n = sample size estimated for the estimate shown as the subscript to n,
k = indicator of level of statistical confidence,
E = acceptable error,
σ^2 = variance of a characteristic in the population,
N = size of the population,
P = proportion of units in the population that have the characteristic,
Q = proportion of units in the population not having the characteristic,
V^2 = relative variance of the characteristic in the population, and
(Rel) = indicates the error, E, has been expressed in percentage form.

It is important to note and keep in mind that these formulae work only for one characteristic and estimate at a time. They are based on the simplest of designs, simple random sampling, and as such produce only rough estimates of sample size and, consequently, costs. Frequently, we will refer back to the formulae throughout the remainder of the paper to show how the terms are defined using information needed from the sponsor as examples. Now consider in more detail the four areas of information required in the survey objectives.

A. Subject of the Survey

A general description of subject matter to be covered by the survey should be provided. However, to clarify the subject the sponsor should specify the major data categories (or subsets) of interest within the general subject. For example, within "labor force statistics," is the sponsor interested in employment, unemployment, minority employment, all of these, or some others? To provide understanding, concepts indicated in the specification of subject matter should be defined clearly. For example, what is meant by "to be employed?" Should the definition include "discouraged workers?" If so, who is considered to be a discouraged worker? As another example, what is meant by "a crime victimization?" When does a fight between husband and wife constitute "assault?"

Once the subject and related concepts have been defined clearly, the information is used to identify target populations. At this stage the information should be provided without having been influenced by perceived administrative limitations.

It is also true that there will be a demand to use some surveys as vehicles for supplemental inquiry. For example, the current CPS design is used in this manner. Obviously, the "best" design would be one established considering also, the need for this supplemental data collection. A complication may exist, though, because the target populations may be different for the basic survey characteristics than the supplemental ones.

Why is information on subject matter needed? A major reason is that the design statistician requires it to be able to compile relevant data required in the determination of sample size. Consider the sample size formulae shown previously. Each of those contains a term (σ^2, PQ, V^2 or Q/P) which was defined as a measure of the variation of a specified characteristic within the "population" of interest. A correctly prepared description of the subject of the survey will specify what that population is as well as populations for which values of other important characteristics are to be estimated. We then have an indication of exactly for which segments of the population data must be compiled. This is described in more detail later in section II.D under "Major Statistics" and "Primary Level at Which Estimate Will be Used."

B. The Purposes of the Survey and the Decisions Based on Survey Results

The next item of importance to the design statistician is the purposes of the survey and the decisions desired. This covers the fundamental reasons why the survey is needed. Frequently, the major reason for a survey may be to help in public policy decision-making and implementation. If so, the statistician needs to know just what the survey is intended to provide for precisely what decisions. An example of this is the use of the State labor force estimates from the Current Population Survey for allocating Federal funds to local areas through the Comprehensive Employment and Training Act (CETA). Here the decision to be made is how much money each CETA prime sponsor will get. The decision depends upon the estimated level of unemployment. Knowing this tells us that a key statistic is unemployment level and, as indicated in section II.D of this paper, we will be concerned with knowing the reliability requirement for the estimate.

The importance of considering the decision-making aspect of the survey is that it allows us to understand more fully the purpose and scope of the survey, and tabulations and types of analysis to be performed. Frequently it can become a basis for determination of the list of key statistics and the reliability requirements for these statistics, as indicated above with the example of CETA funding. It may also help indicate the level of geography, if any, needed in the estimation process, and may aid in determining if the survey is an appropriate vehicle for obtaining the desired information.

C. Types of Analysis and Data Uses

A description of the types of analysis and the uses to be made of the data is also necessary information. This type of information should deal with the specifics of how the data are to be used. It should include a description of tabulations to be made from the survey data and should specify additional types of estimates that might be generated from these tabulations. In conjunction with the estimates, the sponsor should also indicate what tabulations of measures of reliability and accuracy will be needed.

The information should also include a description of what will be done with these data. This should let us know what types of series might be generated, and give an idea of the kinds of adjustments (coverage, seasonal, composite estimation, etc.) desirable to improve estimates. In addition it will help both sponsor and design staff better determine the degree of stratification and clustering desired, help indicate the level for which estimates should be generated and, consequently, help identify which sample sizes need to be estimated.

D. Major Statistics, Their Relative Importance, Primary Level at Which to be Used, and Required Reliability

Of singularly greatest importance to the design statistician, is the area of key statistics and required reliability. Yet, this part of a statement of objectives seldom is established sufficiently early and completely to permit the design statistician to do the best job of initially incorporating the desired design features.

The delay may partially result because the sponsor is not fully aware of what should be provided in this area of the objectives. To be fully sufficient for survey design purposes, a list should be prepared giving: the major statistics desired, i.e., the most important characteristics and the types of estimates to be made of these; the primary levels (e.g., geography, race) at which the estimates will be used; the relative importance (priority ranking) of the key statistics; and the reliability required for each statistic.

It is difficult to discuss each of these independently because of the joint influence they have on final sample size. However, an attempt has been made to do so in the discussion that follows:

Major Statistics

A sponsor can usually specify a large number of tabulations wanted from the data and, generally, how these will be used. The specification of key statistics, though, requires the sponsor to focus on some specific parts of the many tabulations of interest. Thus, in naming key statistics the sponsor is deciding which of many characteristics and estimates (means, totals, rates, changes) best fulfill the purposes of the survey. This is important not only to the design statistician but in many instances such lists have served as a guide for the sponsor to use in reducing the scope of a planned survey when costs prohibit full scale implementation.

It is important that the number of key statistics specified be kept fairly small. Five to fifteen are manageable and will yield sufficient information on sample sizes and other design decisions.

Particular care must be taken in choosing the key statistics. To illustrate, suppose design statisticians were told that some of the estimates of interest were total number of persons having achieved various educational attainment levels (of which "high school graduate" is most important) and total number of employed persons by detailed occupation categories (of which "sales clerk" is most important). To determine sample size required, either equation (4) or (6) would be used, depending upon how the reliability requirement was stated, since totals, not proportions (equation (2)) or means (equation (1)), are specified.

Suppose though that the sponsors had given insufficient thought to specifying the key statistics and that, after the data were collected, the analysis and decision-making processes were only secondarily concerned with total high school graduates and total employed persons that are sales clerks. The main interest turned out to be cross-tabulations of "educational attainment" and "occupation." Since the cross-tab, "high school graduates that are sales clerks," represents a much smaller subgroup than did either of the other two groups, the relative variance would be larger and would have required a considerably larger sample than the one used to get the same required level of reliability as used for the two characteristics separately. In essence, the survey would have failed with respect to its main purpose.

As implied above with the estimates of total, it is also vital to specify which of estimates of level, change over time, and/or comparisons of some type are of prime concern since these specify which of formulae (1) through (6) are to be used in calculating sample size. Sponsors often consider estimates of level of major importance whereas estimates of change or comparisons are considered of equal or lesser importance. In a frequently recurring survey (e.g., monthly or yearly), the level of a statistic is basically determined the first time the survey is conducted. After that, it would seem that the prime interest would switch to estimates of change over time. This is not to say that there is no longer any interest in level or comparisons, but it would seem that the very fact that a survey is conducted repeatedly at relatively short intervals implies prime interest in estimates of change.

In addition to its value in determining sample size, the specification of key statistics is very important in stratification. Knowing the most important characteristics provides the basis for sponsor and design statistician to select those variables (characteristics) which form the "best" stratification criteria.

Primary Level at Which Estimates Will Be Used

The primary level (e.g., geography, race) at which

an estimate is to be used should be provided as part of the definition of the statistic and dictates where reliability measures must be used and should be calculated. As a result, it is also at this level that the sample size estimates are required and must be determined. For example, are estimates to be made and used in analysis and decision-making at the State or national level? Within that geography are the estimates wanted for Blacks and/or Hispanics? Should the estimates be weekly, monthly, or yearly?

Relative Importance (Priority) of Key Statistics

After a list of key statistics is determined, some type of prioritization is needed. This information is important because choices, generally compromises, have to be made in the design. For examples, reliability of one statistic may be improved, but only at the expense of poorer reliability for something else, or due to cost and other constraints, reliability requirements may be achievable for only some of the key statistics. Establishing priorities also helps pinpoint what potential stratification criteria should be given preference for study.

Reliability Required for Key Statistics

So far we have mentioned reliability requirements several times without a definition or discussion of reliability. In general, the reliability of an estimate is a measure of the total error associated with the estimate. In setting reliability requirements in design planning, one should specify the total error that will be considered acceptable under the new design. There are three components which should be considered in setting reliability criteria. These are sampling variance, nonsampling variance, and bias (both sampling and nonsampling).

Sampling variance is the best understood of the three components, and the one that is usually concentrated on in setting reliability requirements. It usually decreases as the size of sample increases, which generally permits the statistician to make acceptable estimates of its magnitude for planning purposes.

For initial rough estimates of sample size, since we have no additional information, we usually assume that sampling variance is the only component of reliability. Under that assumption, the reliability criterion is expressed in terms of the maximum sampling error on the estimate considered acceptable and is expressed as a function both of error (absolute or percentage error) (shown as E in the formulae) and a level of confidence (shown as k in the formulae). For example, a reliability criterion we might be given for estimating "total unemployed" is "No greater than 10 percent error on the monthly sample estimate assuming a 6 percent unemployment rate in the population and a confidence level of 66 2/3 percent."

Given those requirements, we would use equation (6) and calculate $n_{a_{(Rel)}}$ since we are to estimate a total and the desired error has been expressed in relative terms. In that equation the error, E, is 0.10 (the 10 percent error); P is 0.06 "the proportion unemployed"; and k=1 for a 66 2/3 percent confidence level.

With respect to nonsampling variance, many aspects of it are not a direct function of sample size, as is sampling variability. In addition the statistician has considerably less direct information on most of its sources. A number of different phenomena make contributions, e.g., variation over time in the response given by an individual (called response variance) and variance between interviewers.

Response variance is often a particularly large component of nonsampling variances. It is relatively unimportant for simple estimates (e.g., Black unemployment rate), but can lead to serious inference errors in more complex analysis (e.g., determination of the relationship between educational attainment and the unemployment rate).

Bias, like nonsampling error, is contributed by a number of different phenomena and does not decrease as sample size increases; thus, it may be the dominant component of error for very large samples. Nearly all estimates suffer from nontrivial bias. Comparison estimates and change estimates suffer least because biases, at least partially, and sometimes totally, cancel out. Unfortunately, statisticians seldom have good estimates of the overall bias; consequently, bias is usually not fully considered in setting reliability requirements.

Careful questionnaire design and survey implementation, and good quality control can reduce the magnitude of nonsampling variance and bias, but cannot eliminate them entirely. Reliability requirements are important for determining the effort and money that should be put into these activities. For example, if the sole concern in a survey is month-to-month change, bias in estimates of the level of a characteristic may be of little importance. On the other hand, if the prime concern is with level of statistics, for which the sampling variance can be readily made small, then maximum efforts should be made to reduce bias.

Sampling variance, nonsampling variance, and bias can be combined together into a measure called the root mean square error (RMSE), as follows:

$$\text{RMSE} = \sqrt{\text{sampling variance} + \text{nonsampling variance} + (\text{bias})^2}$$

Ideally, reliability requirements should be set in terms of RMSE, and the statistician should design the survey to control the three components of RMSE accordingly. In practice, it is more typical that only sampling variance is considered. This sometimes leads to questionable practices. For example, the sponsor may set a very stringent sampling variance requirement for an estimate of level, and the design of the survey will ultimately achieve it. However, the design statistician may suspect that the bias term dominates the RMSE so that a sample size much smaller than that used would probably have yielded about the same RMSE. Thus, money is apparently wasted or misallocated. Further, if the sponsor really needed a very small RMSE and not just a small sampling variance, the survey as designed has not come close to meeting the actual reliability requirements. In fact, the reliability requirement may not be achievable; thus, perhaps no survey should have been conducted at all.

Setting reliability requirements, even if only in terms of sampling variance, is always difficult. Nevertheless, it must be done. Sometimes the reason for the difficulty is that the sponsor does not fully understand the terminology and

concepts, and thus does not know how to specify required reliability. A role of the statistician is to help the sponsor understand these concepts. Thus, sponsors should not hesitate to seek assistance from the design statistician when there is uncertainty as to the meaning of the concepts.

We believe the key to being able to specify reliability requirements lies in the previously mentioned aspects of survey objectives. If purposes and decisions to be made are specific and well understood, if the desired tabulations and data analyses are well understood, and if the key statistics and their relative priorities are clear, then the sponsor should be able to set at least approximate, minimum reliability requirements for key statistics. In the use of information from survey objectives in calculating initial sample size and cost estimates, we look for statements similar to "Monthly estimates are needed at the State level of the unemployment rate for Blacks with a relative error of 10% or less with a probability of two-thirds."

III. KEY REDESIGN RESEARCH UNDERTAKEN

As mentioned in the introduction, we are redesigning the major current demographic surveys. These are the:

- Current Population Survey (CPS), a monthly labor force survey sponsored by the Bureau of Labor Statistics
- Health Interview Survey (HIS), mostly annual publications, sponsored by the National Center for Health Statistics.
- Annual Housing Survey (AHS), annual publications, sponsored by the Department of Housing and Urban Development.
- National Crime Survey (NCS), mostly annual publications, sponsored by the Bureau of Criminal Justice Statistics.
- Survey of Residential Alterations and Repairs (SORAR), a quarterly survey sponsored by the Census Bureau.

You will recall we stated that every aspect of survey redesign must be related directly to the objectives established for a given survey and that it is the responsibility of the survey design statisticians to consider those in designing the most efficient design possible within imposed constraints. To do that, the information requested from sponsors and users is being considered independently within its respective survey and research is being conducted to determine which procedures of many possible ones are optimal ("best") for that survey.

Because so many different redesign research projects are going forward, we discuss only a few key ones in detail and indicate how the information requested in the first parts of this paper are used in the research or dictate that research be done.

A. Strata and PSU Definitions

Major areas of concern for all the surveys are the formation of strata and primary sampling units (PSU's) within strata. In the past, one set of strata definitions was used for nearly all demographic surveys. However, since the 1970 redesign, there have been three major expansions of the CPS. As a consequence, strata definitions have been modified in a manner which was not anticipated at the time of the post-1970 census redesign resulting in inefficiencies in the current strata definitions. This is a consequence of a shift in the survey objectives from national estimates towards State and substate estimates.

Because of the expanded size of the CPS and the differences between survey objectives for the CPS, NCS, AHS, and HIS, we are investigating alternative stratifications for each survey. In order to accomplish this it is imperative that we have information from the survey sponsors as to what are the key characteristics of interest and their relative importance. This is essential since a cost efficient stratification for one characteristic may not be very good for some of the other characteristics of interest in the survey.

The present PSU definitions were essentially defined about 30 years ago and are the same for all surveys. Except for New England, PSU's are defined in terms of counties. In the redesign, PSU's may vary for the different surveys. We will probably have some subcounty divisions for defining PSU's, especially in some of the counties in the West which cover large land areas and thus involve high interviewer travel costs.

To illustrate what we are doing in defining PSU's, consider plans for the Annual Housing Survey. Some research will be done to determine the optimal work load size per interviewer in each PSU. Most of the research is planned to determine when and how counties and subcounties should be combined to form PSU's. Variance and cost data will be utilized to determine the approximate optimal and upper bound for the geographic size of a PSU. We will then attempt to combine counties in cases where a single county is much smaller than the optimal, and will generally have a one-county PSU whenever the county is larger than the optimal. We are also investigating which set of variables is best in determining those counties to combine to form maximally heterogeneous PSU's. This set of variables will be based on information provided by the sponsor.

B. Rotation Scheme

In most household surveys, we employ a rotation scheme with sample units interviewed several times but eventually rotated out. Because of changes in survey objectives and cost structures, there is a need to reevaluate the rotation schemes now in use.

The CPS has the most complex pattern of any of the surveys, with units in sample for 4 months, out for 8 months, and in for 4 months. This scheme is relatively efficient for estimating month-to-month changes at the national level. However, with changes in survey objectives, there is a great need to get reliable annual estimates of level for State and substate areas as the basis for distributing Comprehensive Employment Training Act (CETA) funds. Thus, we plan to examine rotation patterns that may be more efficient with respect to the variance of annual average estimates for these areas. To that end we will be examining both sampling variation and bias (namely rotation group bias) relevant to each rotation pattern. This examination must be based on the key statistics and the reliability requirements established for those statistics. The final rotation plan decided upon will be that which satisfies the conflicting objectives for State and substate

estimates of annual averages and national estimates of month-to-month change.

The Annual Housing Survey (AHS) has had no sample rotation. Sample units have been interviewed once each year from the time the survey started in 1973. By introducting the redesigned sample in stages we will have the opportunity to determine if there is any rotation group bias in this survey. The results will be used in making decisions about what type of rotation pattern, if any, will be used in the future.

In contrast to the AHS, in the Health Interview Survey each unit is interviewed only once. In order to improve estimates of year-to-year change as well as change over longer periods of time, we will study various rotation plans. There is also considerable interest in obtaining good estimates of annual medical expenditures which would require several interviews at the same housing unit within a year. Because of these specific objectives provided to us by the sponsor, we will consider the possibility of interviewing a subsample of the full sample at about 3-month intervals during the year.

In the National Crime Survey the major concern with the rotation scheme is how it fits in with the reference period. At present, we ask respondents about crime victimizations during the preceding 6 months, and respondents are contacted 7 times at 6-month intervals. Asking respondents about victimizations during the preceding 12 months is an alternative. Then a rotation scheme requring contact with respondents only once a year would be desirable. Here we need to know from the sponsor what is an acceptable level of bias for the various key statistics in order to make a decision among these alternatives.

C. Other Redesign Research

There are many other sample design considerations being studied. A brief description of a few follow.

- All Area Sampling. Area samples are used only in rural areas for all surveys at present, but will probably be used everywhere for the Health Interview Survey because of the sponsor's wishes and confidentiality problems. There is an interest in increasing area list sampling for other surveys as well to simplify sample selection and control. Since area sampling generally results in more variable segment sizes and is more costly than census list sampling, there are potentially bigger gains to be realized by research in this area. Thus, we will be investigating alternative forms of area sampling, trying to minimize cost, variance, bias, and operational problems.
- Telephone Interviewing. A recently completed field study is being analyzed to see if there is any evidence of differences in labor force data between telephone and personal visits for CPS. In the Annual Housing Survey, now conducted entirely by personal interview, a small feasibility test of telephone interviewing will be conducted. In the National Crime Survey, there is a greatly increased use of telephone interviewing this year. The results will be closely analyzed.
- Proxy Respondents. Analysis of a recently completed study is under way to examine differences by respondent type for labor force data. Since past data in the Health Interview Survey has shown important differences between self-respondents and proxy respondents, we are considering a change in the respondent procedure.
- Estimation and Weighting. Research will include improved methodology for noninterview adjustments, imputation techniques, ratio estimation, raking, and composite estimation.

IV. SUMMARY

In the design of new surveys or redesign of the major current demographic surveys, the survey design statisticians at the Bureau of the Census need a considerable amount of specific information regarding sponsor and user needs in order to design the "best" surveys possible within certain constraints such as cost, timing, and/or reliability requirements. The bulk of this information is obtained from the sponsors in the form of a set of specific survey objectives. Since every aspect of survey design and implementation must be related directly to these objectives, it is extremely important that they be clear and thorough. It is also important that they be established early. Several examples were given throughout the paper to illustrate these requirements. Obviously, if a good set of specific objectives are not prepared initially and early-on, the sample size and cost estimates may be faulty, there may be "costly" delays in implementation, and there may be a potential for sacrificing quality for expediency.

BIBLIOGRAPHY

[1] Bailar, Barbara, "The Effects of Rotation Group Bias on Estimates from Panel Surveys." Journal of the American Statistical Association, Vol. 70 (1975), pp. 23-30.

[2] Bailar, Barbara A. and Shapiro, Gary M., "Survey Research at the Bureau of the Census," presented at Symposium on Survey Sampling in Ottawa, 1980.

[3] Cochran, William G. "Sampling Techniques," New York. John Wiley and Sons, (1977).

[4] Gibson, Christina O., Shapiro, Gary M., Murphy, Linda R., and Stanko, Gary J., "Interaction of Survey Questions as it Relates to Interviewer--Respondent Bias," 1978 Proceedings of Section on Survey Research Methods, American Statistical Association, pp. 251-256.

[5] Hansen, Morris H., Hurwitz, William N., and Madow, William G. "Sample Survey Methods and Theory." Vol. I, Methods and Applications." New York. John Wiley and Sons, 1956.

[6] Rubin, J. and Friedman, H. P. "On Some Invariant Criteria for Grouping Data," Journal of the American Statistical Association. Vol. 62 (1967), pp. 1159-1178.

[7] U. S. Department of Commerce, Bureau of the Census. "Sampling Lectures." Washington, D.C. (1968).