

# NONRESPONSE IN FEDERAL DEMOGRAPHIC SURVEYS: 1981-1991

Ayah E. Johnson, Steven L. Botman, Peter Basiotis

Ayah E. Johnson, HMJF 1 Taft Court, Suite 250, Rockville MD 20850<sup>1</sup>

## 1.0 Introduction

### 1.1 Background

The Subcommittee on Nonresponse of the Federal Committee on Statistical Methodology (FCSM) was asked to examine the current state of unit nonresponse in Federal Surveys with specific emphasis on assessing temporal trends in response rates during 1982-1991. This paper focuses on findings only for selected Demographic Surveys.

Concerns about response rates and/or types of nonresponse generally stem from the implications nonresponse has both for data collection and data analysis; and, the knowledge that effective survey designs should incorporate methods to both: (1) minimize the rate of nonresponse while controlling for data collection costs, and (2) employ procedures to compensate/adjust for nonresponse once all practical efforts to minimize nonresponse rates during data collection were exhausted.

### 1.2 Objectives

The main purposes of the study were to assess temporal response rate trends in Federal Surveys from 1981-1991 and to explore factors that could be contributing to the change (if any). While assessing the data we grappled with many of the same issues enumerated and discussed in the CASRO report (1982). Because the data suggested that there was very little or no change in response rates over time, we examined three basic questions:

- What old response rate issues have been resolved?
- What old response rate issues are unchanged? and
- What new response rates issues are raised in federally funded demographic surveys?

### 1.3 Data Analysis

Twenty - six federally sponsored demographic surveys were selected for the study. These surveys were not selected by probability methods because no machine-readable listing of Federal surveys with sufficient auxiliary information was available. Included were Federal surveys conducted either on an ongoing or

on an intermittent basis. Staff of Agencies sponsoring the surveys were asked to complete a questionnaire that was designed and pre-tested by the Subcommittee; this questionnaire was designed to elicit information on nonresponse during 1981-1991, as well as on a variety of survey design features known to possibly affect nonresponse and information on postsurvey adjustment strategies for unit nonresponse.

The Subcommittee itself incurred no unit nonresponse in its data collection activities but incurred some level of item nonresponse. During analysis, issues of measurements, documentation, and dissemination of nonresponse kept surfacing. Item missing data on the questionnaire that the subcommittee fielded were not serious and indicated the kind of data that was not easily reported: (1) Number of hours that interviewers are given to secure response from the sampling unit; (2) Existence of "partial replacement of sample" in successive time periods; (3) Accommodation for proxy respondents who can respond; (4) Number of refusals; and (5) Weighted response rates.

Data analysis, although limited, involved: (1) Examining measurement issues of nonresponse as they compare to those enumerated by the CASRO report; (2) Examining temporal trends; and (3) Identifying survey design features that may affect unit nonresponse.

### 1.4 Study Limitations

Although it is of interest to assess the individual as well as the compounded effects of survey undercoverage, item nonresponse, and unit nonresponse, in this paper the focus is on unit nonresponse only. Both increased resources and use of imputation techniques may have played a role in maintaining the response rate over time. No data collection was attempted for any variable related to data collection costs; anecdotal information indicated that it was not easy to obtain cost, or surrogate cost information for the data collection component in a form that could be related to survey nonresponse rates.

Given the purposive design of the study sample, its small sample size, and the wide variety of survey design differences that characterize these surveys, analysis of these questionnaires should be considered exploratory and treated with caution. It should be recognized that demographic survey samples are typically not selected from highly skewed populations where nonresponse from even a single eligible unit can have a large adverse effect on data quality.

## 2.0 Measurement Issues

The CASRO report in 1982 indicated that although "response rate" designates the ratio of the number of completed interviews divided by the number of eligible units in the sample, its determination depends upon the sample design of the particular survey. Results from the Subcommittee data collection effort

<sup>1</sup> The views expressed in this paper are those of the authors and do not necessarily represent those of their agencies or the Henry M. Jackson Foundation. This paper is based on the work of the Subcommittee on Nonresponse whose members included: Robert M. Groves, Chair, BOC, Fritz Sheuren, IRS, Maria Gonzales, OMB, Dan Kasprzyk, NCES, Susan Ahmed, NCES, J. Donald Allen, NASS, David Belli, BEA, Peter Basiotis, ARS, Steve Botman, NCHS, Eileen Collins, NSF, Mick Couper, BOC, Patricia Guenther, ARS, Paul Hsen, BLS, Ayah E. Johnson, Formerly AHCP, Arthur Kennickell, FRB, Paul Macmahon, IRS, Jeffrey Osmin, BOM, Antoinette Ware Martin, EIA, Pamela Powell-Hill, BOC, Maria Reed, BOC, Carolyn Shettle, NSF.

indicate, as it should, the same result. Rates can be calculated in a multitude of ways, each providing different information that can be used to measure the success of the survey but all depend on the design. The issue is not in the complexity of computing a ratio, but in the expectation that a response rate is one number and that its interpretation is independent of sample design thus allowing for comparison across surveys. For example, in a longitudinal survey with several rounds of data collection, there are several types of response rates that might be of interest: (1) Response rate for each round of data collection; and (2) Response rate for the survey over all rounds of data collection. The same definition holds for both the first and the second response rate, but the computation for the second is slightly more complex than for the first. This implies that when asked to report a response rate for a longitudinal survey, there should be a set of "ratios" that is reported and not a single number.

Thus, although the conceptual idea is to compute a simple ratio as the response rate, the actual calculation should vary with the sample design. It is useful to restate the definitions made in the CASRO report:

**"Response rate:** is a summary measure and should designate the ratio of the number of interviews to the number of eligible units in the sample.

**Completion rate:** is to be considered as a collective term that is used to designate how well a task has been accomplished. In general, completion rates are used to measure how well the various components involved in the sample survey are accomplished."

Other useful measures are employed that may be mistaken for a response rate. The CASRO report in 1982 listed eight such measures, and all are still used in demographic surveys. These measures are helpful in monitoring data collection operations, getting reports of progress from the field, and addressing reactions and difficulties encountered by interviewers. These measures are useful and should continue to be collected. In some instances they approximate a response rate, but they should not be viewed as substitutes for response rates.

Data collected by the Subcommittee indicated that counts of cases by response/nonresponse categories and the distinction between eligible/ ineligible units existed. Twenty - five of the 26 Federal Surveys provided information on sample size, number of ineligible cases, number of interviews, number of nonresponse cases and characteristics of the sample design. The remaining two surveys provided insufficient information to compute the response rates but did provide information on survey characteristics and did provide a response rate. Furthermore, a complete breakdown of refusals and other types of nonresponse was reported only for 14 of the 26 Federal surveys. This was surprising since distinguishing between refusal rates and other reasons for non-interviews is very important. Refusals are less amenable to nonresponse conversion, may require special treatments, and are generally more costly to

convert. Hence there is a need to define and operationalize the concept of refusal (since respondents can refuse to participate in a survey without uttering the words "I refuse"), and there is a need to monitor both the number of refusals and the cost of converting refusals. Whether it is cost-efficient in mean square error reduction to permit interviewers to make large numbers of calls for a respondent contact or nonresponse conversion is questionable (see Groves 1989).

There are several reasons why the count of cases by sources of nonresponse was missing. Not all traditional sources of nonresponse such as refusals or "not at home" are applicable for all demographic surveys. For example there are no refusals for surveys that extract data from administrative records. Other responses were missing because the data were not readily accessible for reporting, although they were monitored. Even though counts were reported by response/nonresponse categories, *there was still no uniformity in computation of response rates.*

One additional issue that arose was the interchangeable use of two statistical concepts, "sampling units" and "analytical unit". In element sampling, the ultimate sampling unit contains one element, whereas in cluster sampling, the ultimate sampling unit may contain more than one element. After data processing, the "element" of interest is called an "analytical unit". Response rates are generally computed at the level of the sampling unit. For example (hypothetical), in a survey of schools the sampling unit is a school, whereas the analytical unit could be the school, the teachers within the school, or the student population within the school. When analyzing survey responses from teachers, the response rate of interest is the proportion of teachers that were eligible and responding however what may be reported is the proportion of schools that were eligible and responding. Irrespective of the "unit," data on rates by major survey characteristics were not easily reported.

Most demographic surveys in this study were not based on self-weighting designs. If all elements in the sample were equally likely to be selected (a self-weighting sample), the unweighted and the weighted response rates would be the same. In such a case the unweighted response rates can provide both the required measure of sample representativeness and the measure of success by field operations in securing a response. If, however, the elements of the population that are being selected are disproportionately sampled, the unweighted response rate can only provide one overall measure of quality of field operations, whereas the weighted response rates provide a measure of the representativeness of the population. These weighted response rates in a sample survey are essential to ascertain the representativeness of survey data and to assess the effect of nonresponse on estimates of interest. All demographic surveys provided a response rate, but it was not always weighted. Therefore, there is no

evidence of the comparability of the response rates for a broad array of Federal surveys. However, for most demographic surveys the unweighted response rate approximates the weighted response rate. The amount of the difference depends on the variability of the probability of selection for the survey and whether nonresponse propensity is related to the probability of selection.

Response rates reported for these surveys were computed either during data collection (about 52 percent) or during data processing (about 48 percent). Some surveys reported response rates<sup>2</sup> and others provided completion rates or proportion of the sample that was interviewed -- usually for the sampling units.

To summarize, the general guidelines given in the CASRO report appeared sufficient to stimulate the collection of nonresponse data but were not sufficient to achieve uniformity and comparability for computation of response rates. The data collected by the Subcommittee on Unit Nonresponse demonstrated that information concerning the sample design, field procedures, survey characteristics, and methods for adjusting for nonresponse can be collected. The next step may be to prepare and disseminate guidelines for computing and reporting response rates that are sample design specific.

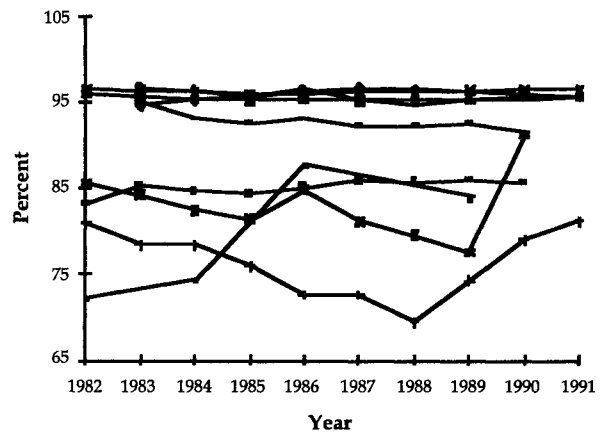
### 3.0 Data Analysis

#### 3.1 Temporal Trends

Analysis of response rates over time was restricted to those surveys with at least four points for survey data collection during 1982-1991 (no temporal data for 1981). Only 8 of the 26 demographic surveys included in the data collection met this criterion. In this analysis we examine: (1) The calculated response rate for these 8 demographic surveys; (2) the mean nonresponse rates for demographic surveys broken down by percent of refusals and percent of non-contact; and then (3) focus on two specific ongoing surveys, the National Health Interview Survey (NHIS) sponsored by the National Center for Health Statistics (NCHS) and the Current Population Survey (CPS) sponsored by the Bureau of Labor statistics. All presented response rates are unweighted and calculated based on counts provided by the respondents.

Figure 1 displays the calculated response rates over time for the eight demographic surveys. Although no evidence was found to support a downward trend on survey response rates among Federal demographic surveys from 1982 to 1991, surveys that were conducted on a yearly basis exhibit smaller fluctuations in their response rates than those that were conducted less frequently.

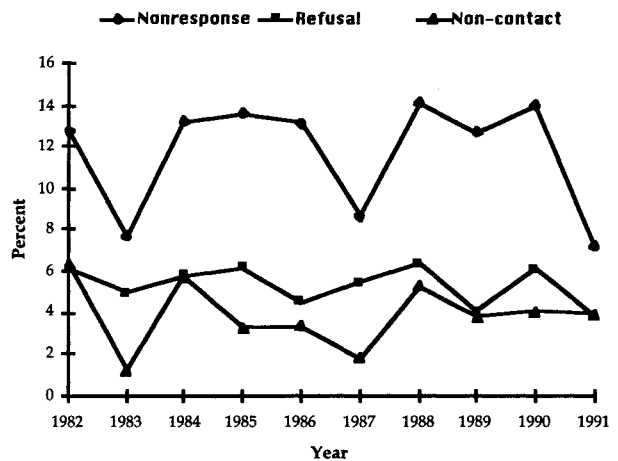
Figure 1. Temporal Trends



This fluctuation can be explained by separating the studies into two groups. One group had response rates in the 95-percent range, and the second cluster was about 10 to 15 percentage points lower. The studies in the 95-percent range consisted of ongoing studies, often with panel components, conducted by the same interviewer corps. In addition, refusal rates for surveys conducted on a yearly basis were almost half those of the less frequent surveys. Neither group exhibited a strong consistent trend over time.

Figure 2 displays the mean nonresponse rates for demographic surveys. It also shows the breakdown by two types of nonresponse-- refusals and non-contact. In 1983 and in 1987 there was a down-trend in both overall nonresponse and non-contact rates while the refusal rates declined in 1986 and in 1989. The refusal rates seem to be more stable over time than the non-contact rates. One can speculate that although it is hard to gain cooperation from reluctant respondents, at least we can locate them and enlist their cooperation; others are difficult to contact at all.

Figure 2: Mean Nonresponse Rates



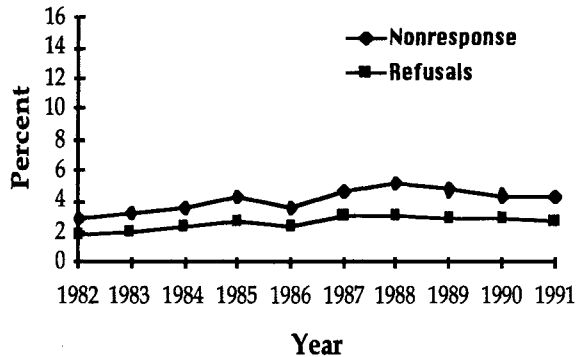
#### 3.2 The CPS and the NHIS

Two major surveys collected data and reported

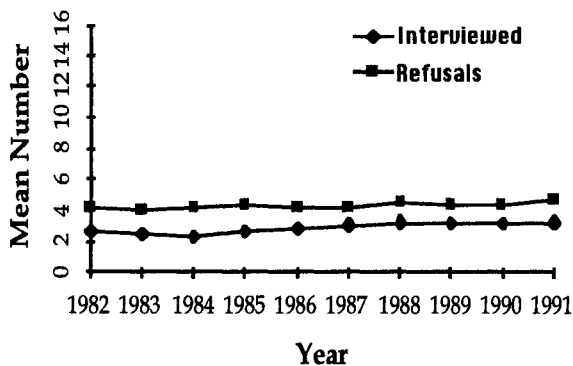
<sup>2</sup> A response rate is the ratio of the number of responding units to the number of eligible units; a completion rate is the ratio of number completed to number fielded; and the proportion of number interviewed is the ratio of those interviewed to those fielded.

response and nonresponse rates for 1982-1991: The NHIS and the CPS Temporal trends are depicted for the NHIS and the CPS in Figures 3 and 5, respectively. Moreover, for the NHIS we were able to obtain information on the average number of call-backs for both completed interviews and refusals-- Figure 4.

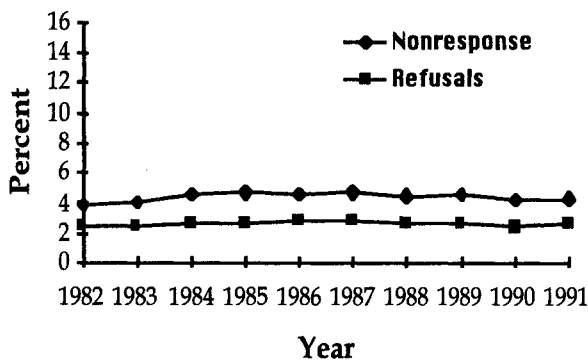
**Figure 3. Nonresponse & Refusal Rates-- NHIS**



**Figure 4. Mean Number of Calls Per-Household--NHIS**



**Figure 5. Nonresponse & Refusal Rates-CPS**



Over the decade response rates for neither the NHIS nor the CPS changed dramatically. For the NHIS, the line depicting the refusal rate almost parallels the plot for the nonresponse rate. The refusal rates constituted, between 61 and 68 percent of the overall

total nonresponse. For the CPS, the pattern was similar. The refusal rate in this case constituted between 57 and 64 percent of total nonresponse. However the CPS temporal trend from 1955 to 1990, showed an increase in the refusal rate (Tucker, work in progress).

This also indicates that response rates were not significantly changing. At least two explanations can be offered: (1) Response rates have been stable but they are costing more to maintain; and (2) extensive use of imputation techniques can allow for a less restrictive definition of a responding unit.

Although the stability of the response rate for the NHIS was consistent over the years, the average number of call attempts for interviewed cases increased (see also Kalsbeek et al., 1994). On average, 2.7 calls were required to complete an interview in 1982, compared with 3.2 calls in 1991. The increase was more pronounced for calls dealing with refusal rates. There were on average 4.1 calls for handling refusals in 1982, compared with 4.7 calls in 1991. Thus, the mean number of calls provided evidence that increased efforts were required to maintain response rates in the field.

### 3.3 Survey Design Features

In addition to examining response rate trends over time, the Subcommittee examined how response rates may vary across design features.

**Survey Frequency.** The response rates for all surveys over all years ranged between 67 and 96 percent (Figure 1). Demographic surveys clustered into two major groups: (1) those that were conducted at least four times during 1982- 1991<sup>3</sup>; and, (2) those that were conducted less than four times during the same period. When analyzing the response rates over time one can see two distinct bands. Although the differential between the two bands of surveys was about 4 percent, no statistical difference was detected between response rates for the more frequent and the less frequent surveys. The main differences detected were in refusal rates. Refusal rates for the less frequent surveys were almost twice those reported by more frequent surveys.

One of the working hypotheses was that less frequent surveys may use a more complex and time-consuming core questionnaire or deal with more difficult topics. As part of this study we collected information on the amount of time it takes to complete the core questionnaire. A cross tabulation of survey frequency and time needed to complete the core questionnaire revealed that this was not the case. For 33 percent of the less frequent surveys it took more than one hour to administer the core questionnaire, compared with 27 percent for the more frequent surveys. This may indicate that the length of the core questionnaire reflected analysis objectives rather than survey frequency or that

<sup>3</sup> The definition of "frequent surveys" as ones which were fielded at least four times during the ten - years reference period, was based on the distribution of the surveys' "frequency".

most on-going surveys had lower refusal rates because they employed more experienced interviewers who had experience in obtaining the required cooperation. For the less frequent surveys there was a learning curve for interviewers and for all involved in fielding the survey.

**Length of Interview.** This design factor was consistently inversely related to response rate-- the longer the interview, the higher the response rate. This result reinforces the notion that once at the door, the quality and not the length of the instrument will play a major role in obtaining a response.

**Sampling units.** Five surveys sampled the "household" and interviewed one or all persons in the "households". 17 surveys sampled and interviewed persons. The remaining 3 surveys consisted of physician and school surveys. On average the response rates are higher for household surveys (90 percent), followed by surveys of persons (82 percent) and finally other surveys (82 percent). Refusal rates were comparable for the first two groups of surveys, households and persons, and slightly lower for the other surveys.

**Data Collection Agent.** We distinguished among surveys whose data collection was conducted by Federal agencies, academic institutions and contract organizations working for the Federal agencies. The mean response rate for surveys conducted by Federal agencies was 88 percent, by contract organization, 79 percent and by academic organizations 76 percent. This differential may reflect the frequency and the difficulty of the survey or the resources available to the different organizations.

**Mode of Data Collection.** Most demographic surveys collected their data using more than one mode of data collection. Eleven of the 26 surveys used face-to-face interviews as the main mode of data collection with additional contacts made by telephone, mail, or use of administrative records. Five of the surveys were conducted using only face-to-face interviews. One survey was conducted only by phone, and seven others have a combination of telephone, mail and extraction from administrative records. The remaining two were based only on administrative records or the "other" group. It is clear from this distribution that most surveys try to establish some verbal communications with the respondents and used more than one mode of data collection to try to reduce the rate of nonresponse.

**Number of supplements.** Number of supplements administered did not seem to affect the response rate but they did affect the refusal rates. The existence of supplements indicated a higher refusal rate. Also, there were problems in defining what should be counted as a supplement, and no measure on the length of the supplement was available.

### 3.4 Postsurvey Adjustments

Eleven of the demographic surveys used poststratification, 22 used ratio adjustments (weighting up), 9 used raking, 5 used regression and one used imputations. The common factor present in all these

demographic surveys was that Federal agencies use one or more forms of postsurvey adjustments. This area could be considered one where changes were being made, and analytically each agency was trying to minimize or adjust for potential nonresponse bias.

Use of postsurvey adjustments may also have been a contributor to the tolerated level of response rates. Since there is methodology that could be used to handle issues of item nonresponse, we do not know whether the classification of unit nonresponse changed on a survey-by-survey basis. Thus a new issue not previously reported, involves the classification of when a case should be considered a response. Computation of response rates may be tangled with the issue of imputation for item nonresponse.

### 3.5 Documenting Response Rate

All 26 demographic surveys maintained some information about response/ nonresponse components. Fifty-nine percent of the demographic surveys tracked five or more different components simultaneously. The component most frequently documented was "refusals" (14 of the responding surveys). Other frequently recorded components were "temporarily absent" (10 surveys), "not at home" (10 surveys), "ineligible" (8 surveys), and "language barriers" (7 surveys). In addition, the number of cases for each category of response disposition was generally available for the same number of years as the overall response rates, although, over the decade there was some variation in which components were recorded. Response rates classified by main demographic characteristics were either not tracked or harder to obtain since they required intensive computer manipulation of data bases.

## 4.0 Conclusions/Summary

Despite the study's focus on nonresponse rates, major difficulties arose in getting consistent information on response rates. Computed rates, identified as "response rates," have different names and different definitions depending on the surveys and when they are collected (for example, during data collection or during report writing). In that sense the measurement issues outlined in the CASRO report have not changed.

Reporting practices and documenting response-rate components varied widely across surveys. Demographic surveys maintained information about response/non-response components but not in an easily accessible data base. Most surveys used definitions and concepts that were tailored for their specific needs.

Temporal trends did not seem to indicate a decline in response rates although for some of the demographic surveys the non-contact rate fluctuated. Refusals seemed to be stable-- there was a core of persons or institutions that refused to participate.

Post--survey adjustments used to reduce the effect of nonresponse were: Poststratification, ratio adjustment, raking, regression modeling of the propensity to respond, and imputation. Some of the approaches were traditional while others were at the

cutting edge of best practice.

One key issue that needs to be addressed in future studies is the cost (not necessarily in dollars) to maintain these response rates. Variables pertaining to call-back rules, expectation on nonresponse rates, mean square error, as well as cost per case could shed light on reasons for higher or lower response rates.

Another issue to be addressed is the impact of the use of CAPI on both unit nonresponse and item nonresponse. Early results from various users of a CAPI indicate a lower rate of item missing data; as far as unit nonresponse is concerned, it is not clear whether the use of CAPI has an impact. Thus with new technologies being used for data collection, the nonresponse issues are likely to be different.

## 5.0 Recommendations

Recommendations of the Panel on Incomplete Data (Madow et al, 1983) are still valid, and judging from the information that was collected many have not been implemented. We review the status of a few. Recommendation # 4 of the panel on incomplete data was to "*compute nonresponse rates during as well as after data collection, for important domains, and for important items*". This recommendation has been partially implemented. Federal agencies compute rates, but those are not necessarily response rates and, for the most part, they are not broken down by domains. Moreover, for Federal demographic surveys these numbers are not part of an information system that is easily accessible.

One can speculate about the reasons for having multiple definitions of rates and for not having a system that track response rates for a survey both at a given point and over time. First the recommendation, although clearly stated, did not elaborate on how this should be done, and unlike the CASRO report the subcommittee did not advocate uniformity in definition.

A second reason is associated with resources, costs and benefits associated with such a monitoring system. Researchers are interested in the data that has been collected and not the data that has not been collected.

A third reason is quite simple. If the interviewers encountered unit nonresponse, then the domain to which the unit belongs is unknown. We would need to devise a mechanism to capture some data on nonrespondents.

The recommendations from this study are to support the panel's recommendation with guidelines for computation of response rates that will be design specific and will provide the necessary information for computing and tracking response rates over time. Also, since controlling for nonresponse starts with the design and the data collection phase, these guidelines should be extended to rates other than nonresponse that will be helpful in monitoring and reducing nonresponse, when it is encountered. In order to compute domain-specific rates a strategy for estimating domain specific nonresponse rates must be adopted -- a subsample of nonrespondents can be part of the design.

*Survey staff should monitor response rate components over time in conjunction with routine documentation of costs and design changes.* Although response rates are one of many measures of data quality, they are useful tools in monitoring changes in the quality of survey statistics. For repeated surveys, a time series of response-rate components, juxtaposed with costs for each wave and indicators of design changes introduced in any wave can be valuable management tools. Ideally, nonresponse components should be presented for major analytic subgroups. This would provide the consumer of statistical reports with consistent information about nonresponse properties of the statistics.

*A centralized data base of response rates and survey design features should be constructed to help explore the influences on magnitudes of response rate components.* The Subcommittee's effort was focused on a small number of surveys but required hundreds of hours of work by its members and survey staffs -- all to assemble information that is critical in assessing the ability to disseminate information. Further, by including key design features of the surveys, observational studies of correlates of response rates can be made. Such a data base would be useful in identifying current and temporal trends in survey response rates. The current international effort at compiling this information (see De Heer and Israels, 1992) is compatible with this recommendation. If the international effort at creation of such data bases is successful, the US. will have comparative data cross-nationally.

*Full sample data sets should be given in public use data files.* The Subcommittee found that this was the exception not the rule in Federally funded surveys. Public Use Data sets include only the respondent data file. Releasing a complete data set with the selection weights, allows the analysts to construct alternative postsurvey adjustments for the nonresponse.

## REFERENCES

- Kalsbeek W.D., Botman S.L., Massey J.T., & Pao-Wen Liu, 1994.* Cost-efficiency and the Number of Allowable Call Attempts in the National Health Interview Survey. *JOS*, Vol 10, No. 2: 133-145.
- Steeth Charlotte G.,1981.* Trends in Nonresponse Rates 1952-1979. *Public Opinion Quarterly* Vol. 45.
- CASRO, 1982.* On the Definition of Response Rates. A special Report of the CASRO Task force on Completion Rates, Lester R. Frankel, and Published by the Council of American Survey Research Organizations.
- Groves R.M., 1989* Survey Errors and Survey Costs. John Wiley and Sons, Inc. NY 1989.
- Madow et al., 1983.* Incomplete Data in Sample Survey, Vol I.
- Subcommittee on Survey Coverage, FCSM,1990* Statistical Policy Working Paper 17, Office of Management and Budget, Washington DC.