Survey Nonresponse: New Definitions and Measurement Methods

Nancy Bates, Pat Doyle and Franklin Winters[1]
U.S. Census Bureau

---

1. Standards for Measuring Nonresponse

> "Statistical surveys are measurement processes and, as such, need to use reproducible methods.   In order to produce data that can be used for any purpose, but especially for policy purposes, we need surveys that are in statistical control."
>
> - Barbara Bailar (1987)

Information about federal survey nonresponse provides critical performance measures for all aspects of the survey process. Nonresponse information helps the data collection agency identify areas, items, and questions that need improvement, informs data users of the quality of unit and item response, and helps field managers evaluate the performance of employees. Ideally, nonresponse measures should also serve as quality benchmarks across different surveys. But to do so, nonresponse measures must adhere to some degree of consistency in the concepts and definitions that comprise the given measure. For example, what do we mean by an 'eligible' sample unit? What defines a case as 'out-of-scope'? Unfortunately, there are as many answers to these questions as there are federal surveys currently being fielded.

The attempt to document nonresponse measures and establish conformity across the survey community is certainly not a new idea. In 1977, the Committee on National Statistics published several volumes dealing with incomplete data in sample surveys (Madow et al, 1983). Likewise, the Council of American Survey Organizations (CASRO) issued a special report on response rate definitions in the early eighties (CASRO, 1982). A decade later, the Office of Management and Budget sponsored a Federal Committee on Statistical Methodology subcommittee to document and study trends in unit nonresponse and the measures used to compute them (Shettle et al., 1994; Johnson et al., 1994).   The FCSM study found little consistency among the reviewed surveys in how they measured and reported nonresponse rates. These inconsistencies primarily stemmed from differences in sample design across the surveys.

Even within agencies, the concept of a standard nonresponse rate is blurred. For example, in 1986, the survey implementation branches of the Census Bureau took inventory of how each area defined various nonresponse situations. Later, methodologists from each area reviewed these results and discovered that differences were obvious.   Reasons for differential treatment of nonresponse were attributed to several factors including: lack of resources, tradition, isolation, and legitimate design differences (Bailar, 1987).

Given the less-than-successful attempts to previously establish standard definitions and reduce differences in nonresponse calculations, why try again? One answer is simple: nonresponse rates to federal surveys are increasing.   Simply put, this means the potential for nonresponse bias is also increasing with negative consequences to both the data collection and data analysis. If survey managers hope to understand the reasons behind the increase, we must first have accurate measures to reflect the extent of the problem. Since no two surveys are exactly alike in their purpose, sample design, content, interview period, mode, respondent rules and periodicity, the federal statistical community needs a core set of standard nonresponse measures that minimize design differences and can be applied across different surveys. Additionally, we need to establish standard definitions beyond unit nonresponse that provide survey-specific insight to

nonresponse (e.g., attrition rates in longitudinal surveys, item nonresponse, person-level nonresponse).

A review of the literature quickly yields many recommendations for the types of nonresponse measures federal statistical agencies *should* be producing as well as suggestions for how to go about it. For example, in the Committee on National Statistic's *Principles and Practices for a Federal Statistical Agency*, the committee recommends as two of their key practices "openness about the data provided" and "coordination with other statistical agencies" (Martin and Straf, eds.; 1992). To this end, the committee recommends providing reliable indicators of the amount and types of error survey estimates are subject to. Quality profiles that contain more than just estimates of sampling error are cited as one useful example. Similarly, in *Best Practices for Survey and Public Opinion Research*, the American Association for Public Opinion Research (AAPOR) recommends an honest and objective full reporting of nonresponse measures (both unit and item nonresponse) along with documentation and full description of response and completion rates (AAPOR, 1997).

In the UK, the statistical quality checklist from the Office for National Statistics (ONS) suggests provision of tables showing whether nonresponse is due to non-contact or refusal as well as item nonresponse measures for key items (ONS, 1997). More recently, the ONS issued a paper specific to the standardization of response rate estimation for social surveys (Lynn, Laiho, Martin and Beerten, 2000). In the US, the AAPOR released *Standard Definitions* (1998) outlining case code definitions and formulas for calculating response rates for random digit dial (RDD) telephone and in-person surveys. This publication provides detailed guidelines for calculating response rates, cooperation rates, refusal rates and contact rates. Finally, the FCSM recently established a subcommittee to promote wide-spread development of quality profiles across government agencies and surveys. The committee is currently finalizing their report which will include a recommendation to document levels of unit nonresponse error in surveys along with a discussion of several formulas used to produce them (OMB, 2000). In summary, professional associations and oversight committees alike have gone on record that survey organizations should produce consistent and comprehensive nonresponse measures – meeting this challenge within the federal survey community is the focus of this paper.

In the discussion that follows, we describe the recent efforts of an interagency group charged with implementing this broad recommendation. We begin by briefly describing the group's experience in tracking unit nonresponse trends over the decade for a select number of continuing demographic surveys. We then discuss the group's involvement in defining a core set of unit nonresponse measures that were calculated for the same set of surveys. Next, we describe recent efforts to define a core set of survey-specific nonresponse measures. Finally, we conclude with a discussion of how the Census Bureau plans to get these nonresponse measures calculated and published on a routine basis in a standardized report.

2. The IHSNG and its role in devising core nonresponse measures

It would be misleading to suggest that federal statistical agencies never calculate or publish measures of survey nonresponse. In fact, they do so regularly in reports such as the CPS monthly summary report and other status reports generated for survey sponsors. Histories of yearly response rates are published less frequently in documents such a survey quality profiles and

technical papers (see U.S. Department of Commerce, 1998; U.S. Department of Labor, 2000). But the fact remains that official publication of many response rates remains fragmented, sporadic, and non-standardized.

In the Spring of 1997, the Census Bureau and several of its sponsoring agencies formed a new interagency committee to address current nonresponse issues. As its first project, the group updated and reviewed nonresponse rates since the last interagency group had done so some 10 years earlier. The committee (known as the Interagency Household Survey Nonresponse Group or IHSNG) focused its efforts on six large, continuing household surveys. These included: the Current Population Survey(CPS), the Consumer Expenditure Quarterly (CEQ), the Consumer Expenditure Diary (CED), the National Crime Victimization Survey (NCVS), the Survey of Income and Program Participation (SIPP) and the National Health Interview Survey (NHIS). The IHSNG was successful in documenting annual average nonresponse rates between 1990-1997 for these surveys but in the process discovered that the surveys lacked consistent definitions of nonresponse concepts and statistics. Consequently, the group recommended that a core set of nonresponse statistics be identified for comparison purposes across surveys having different designs. They also recommended developing an expanded set of consistent nonresponse measures appropriate to the goals of specific surveys (Atrostic and Burt, 1998). In 1999, staff from the Census Bureau, the Bureau of Labor Statistics (BLS), and the National Center for Health Statistics (NCHS) formed IHSNG subgroups to implement these recommendations.

One subgroup approached the first recommendation by again focusing on the six continuing surveys studied previously. Admittedly, this eased the task for several reasons. First, although the sponsors differ, data for each of the surveys is collected by only one agency – the Census Bureau. Second, although many of the design aspects differ, the *initial* contact for each survey is by personal visit. The group gathered the definitions and exact formulas used to generate the average annual nonresponse rates previously produced. Group members debated alternative definitions but ultimately agreed that the most appropriate 'generic' or standardized measures of nonresponse across surveys having different features are the interview rates based upon initial contact, or in other words, the first time a unit is in sample. The group reasoned that nonresponse rates measured at the initial interview can be more readily applied across different surveys because they control for several of the design factors hypothesized to influence nonresponse such as number of interviews, frequency of interviews, interview length and, in this case, mode of interview. It is important to note , however, that design differences still influence response rates and should be considered when interpreting rates based on initial contact (e.g., length of field period and survey subject matter).

Having agreed upon this general concept, the group took inventory of the survey outcome codes retained at the point the Census Bureau calculates nonresponse rates. They discovered that the surveys currently gather similar information and define categories of nonresponse and eligible units in a fairly similar manner. Hence, the group recommended a short term approach to compiling comparable trends in nonresponse rates by concentrating on the categories currently available. They learned that the surveys currently distinguish interview outcomes and categorize cases as noninterviews according to the following table:

Table 1.
Noninterview Reasons Currently Captured in Selected Surveys

Survey	Refused	No One Home	Temporarily Absent	Language Problem	Other Reason
CEQ	X	X	X	—	X
CED	X	X	X	—	X
CPS	X	X	X	—	X
NCVS[1]	X	X	X	—	X
NHIS[2]	X	X	X	X	X
SIPP[2]	X	X	X	X	X

1- Breakouts of refused, no one home, temporarily absent and 'other' available starting in 1994.
2- Breakout for language problem available starting in 1995 for NHIS; 1996 for SIPP.

3. Core Set of Unit Nonresponse Measures

Using these outcome categories as a template, the group defined seven common rates that could be calculated by most of the six surveys as a way to decompose nonresponse in the short term. The core set of nonresponse measures developed by the IHSNG subgroup and subsequently validated by staff from the Census Bureau, National Center for Health Statistics, and the Bureau of Labor Statistics include:

· The Initial Interview Response Rate (IIRR)
· The Initial Noninterview Rate (INR)
· The Initial Noninterview Language Problem Rate (INLR)
· The Initial Noninterview Temporary Absence Rate (INTAR)
· The Initial Noninterview Not at Home Rate (INNHR)
· The Initial Noninterview Refusal Rate (INREFR)
· The Initial Noninterview Other Reasons Rate (INOR)

Before providing the exact variables, definitions, and formulas for calculating these rates, it is critical to emphasize four assumptions:

·   Assumption #1: The rates apply only to the first time a unit is in sample. For panel surveys that follow addresses, this means using the outcome at the end of the first interviewing cycle. For panel surveys that follow people, this means using units in the first round or wave of interviewing. For one-time, annual surveys, this means using units in sample at a specified time.   When calculating an annual initial nonresponse rate for a monthly survey with rotating samples (like the CPS), the correct computation is to sum all of the initial contact noninterview types relevant to that particular rate (e.g., all first contact noninterviews due to language problems), over the twelve months in sample and divide by the sum of all initial contact eligible units over the twelve months in sample.

·   Assumption #2: Sample units with undetermined eligibility status are to be considered eligible and placed in the numerator and denominator when calculating the initial contact nonresponse rate. (We note that all initial contacts for the six surveys studied here are in-person. We acknowledge that assumption #2 should be reconsidered in surveys where the initial contact is by phone, particularly in random-digit-dial surveys).

·   Assumption #3: When available, the rates should be calculated using the final, edited outcome codes. For more recent years when final codes may not be readily available, the interim outcomes may be used but must be noted in the text and/or tables when presented.

·   Assumption #4: For surveys that undergo post-delivery sponsor edits that result in changes to interview outcomes, the sponsoring agency will be the source for the initial contact nonresponse data, otherwise the collecting agency will be the data source.

Variables Necessary to Compute Core Nonresponse Rates

Variable   **NEU**: Number of Eligible Units
Definition   Number of interviewing units in the sample that are considered to be eligible for interview, e.g., they are determined to be existing structures that are residentially occupied. This category also includes units of undetermined eligibility (see assumption #2).
Computation   Sum of all units classified as eligible.
Details   Definition of 'eligible' may vary across surveys.

Variable   **NINT:** Number of Interviewed Units.
Definition   All interviews considered by a predetermined definition to be complete.
Computation   Sum of all interviewed units.

Variable   **NENIU:** Number of Eligible Non-Interviewed Households.
Definition   The sum of interviewing units eligible for interview that were not interviewed because of: language problems, no one home, temporarily absent, refusals, and all other reasons.

Computation   Sum of all eligible units classified as noninterviews.

Variable   **NEUE**: Number of Eligible non-interviewed Units Excluding refusals, language problems, no one at home, temporarily absent.

Definition   Interviewing units that are eligible for interview but are not interviewed for some reason other than a refusal, a language problem, no one being home, or the interview unit being temporarily vacant. This category includes non-interviewed units with undetermined eligibility (see assumption #2).

Computation   The sum of all eligible units classified as noninterviews for reasons other than language problems, no one home, temporarily absent or refusal.

Details   Some surveys do not capture the language problem separately and these cases fall into this category (e.g., the CPS and the CE and NHIS before 1995).

Variable   **NREF:** Number of Eligible Units Refusing to be Interviewed**.**

Definition   Interviewing units eligible for interview where the respondent refuses to be interviewed .

Computation   Sum of cases classified in the NREF category.

Variable   **NEUNH**: Number of Eligible Units not interviewed due to No One Home.

Definition   Number of interviewing units never interviewed because no one was ever found at home within the interviewing period.

Computation   Sum of cases classified in the NEUNH category.

Variable   **NEUTA:** Number of Eligible Units not interviewed due to Temporary Absence.

Definition   Number of eligible interviewing units never interviewed because occupants are away temporarily (e.g., on vacation) during the field interviewing period.

Computation   Sum of cases classified in the NEUTA category.

Variable   **NEUL:** Number of Eligible Units not interviewed due to Language problems.

Definition   Interviewing units eligible for an interview where an interview was not obtained because the respondent could not converse in the language of the interviewer or available translator.

Computation   Sum of cases classified in the NEUL category.

Details   Some surveys do not currently capture the language problem separately (e.g., the CPS and the CE and NHIS before 1995).

Core Unit Nonresponse Rates and Formulas

Rate   **IIRR**: Initial Interview Response Rate

Definition   Number of interviewed interviewing units divided by the number of eligible interviewing units.

Computation   (NINT/NEU) x100

Rate   **INR:** Initial Noninterview Rate

Definition   Combination of eligible interviewing units that were not interviewed due to language problems, refusal, no one home, temporarily absent or other reasons divided by the total number of eligible interviewing units.

Computation   (NENIU/NEU) x 100

Rate   **INLR:** Initial Noninterview Language Problem Rate

Definition   Number of eligible interviewing units not interviewed because of language problems divided by the total number of eligible interviewing units.

Computation   (NEUL/NEU) x 100

Rate   **INTAR**: Initial Noninterview Temporary Absence Rate

Definition   Number of eligible interviewing units not interviewed because occupants were temporarily away during the interview period divided by the total number of eligible interviewing units.

Computation   (NEUTA/NEU) x 100

Rate   **INNHR**: Initial Noninterview No One Home Rate

Definition   Number of eligible interviewing units not interviewed because occupants were never found to be at home during the interview period divided by the total number of eligible interviewing units.

Computation   (NEUNH/NEU) x 100

Rate   **INREFR**: Initial Noninterview Refusal Rate

Definition   Number of eligible interviewing units not interviewed because occupants refused to participate divided by the total number of eligible interviewing units.

Computation   (NREF/NEU) x 100

Rate   **INOAR:** Initial Noninterview Other Type A Rate

Definition   Number of eligible noninterviewed interviewing units excluding refusals, language problems, no one at home, temporarily absent divided by the total number of eligible interviewing units.

Computation   (NEUE/NEU) x 100

4.1 Comparison between IHSNG and AAPOR Definitions

As mentioned previously, in 1998 the AAPOR published standard definitions with documentation of rules for calculating survey response rates. These were developed as a public service to the survey research industry and are based upon a committee that reflects both public and private survey organizations. Consequently, it is useful to present a comparison between the IHSNG definitions and the AAPOR's -- in doing so, we hope to highlight the fact that the two are complementary in many ways, thus avoiding confusion about which definition is most appropriate to use when calculating a response rate for comparison purposes.

As noted in Table 1, the IHSNG core nonresponse measures are driven in large part by the level of detail currently captured by Census Bureau final outcome codes. When comparing these codes to the final disposition codes recommended by AAPOR, we find several differences. Some differences are simply due to varying degrees of detail while others stem from the AAPOR inclusion of RDD telephone surveys (whereas ours are limited to in-person interviews since initial contact for the surveys studied are all personal visit) . For example, the AAPOR recommends four broad categories of outcome that cover both RDD and in-person, household surveys. These include: 1) Interview, 2) Eligible, non-interview, 3) Unknown eligibility, non-interview, and 4) Not eligible. Complete interviews and partial interviews are sublisted by AAPOR under the broader 'interview' category. Both situations are equivalently collapsed under the 'complete' code using the Census outcome capture scheme (Census records interim outcomes of partial or sufficient partials but these are converted to 'complete' outcome codes in the final disposition for some surveys). Using this as an example, we see the level of detail is less with the Census outcome code in some cases, yet the underlying components that comprise the larger category should be equivalent to that of AAPOR.

Under the eligible, non-interview category, the AAPOR lists three main nonresponse subcategories for cases where no interview is obtained. These include 1) refusals and break-offs, 2) non-contacts, and 3) other. The IHSNG core nonresponse rates are based on five nonresponse categories (refusals, no one home, temporarily absent, language problem, and other) as opposed to three, but the no one home and temporarily absent groups both fall under the AAPOR non-contact category and the language problem category falls under AAPOR's 'other' group. In this example, the Census Bureau's outcome codes actually allow us to go beyond the level of response rate detail outlined in the AAPOR guidelines.

Additionally, the core nonresponse formulas share many similarities with the formulas published by AAPOR. The Initial Interview Response Rate (IIRR) for example, is very similar to AAPOR's Response Rate 2 (RR2). The RR2 counts complete and partial interviews as respondents and places all cases of unknown eligibility in the denominator (AAPOR 1998, pg. 18). Similarly, the IIRR places both completed and sufficient partial interviews in the numerator and includes all eligible cases in the denominator (including those of unknown eligibility -- see assumption #2). Likewise, the Initial Interview Refusal Rate (INREFR) is very similar to AAPOR's Refusal Rate 1 (REF1). The REF1 is the number of refusals divided by interviews, non-interviews, and cases of unknown eligibility (AAPOR 1998, pg. 21). The INREFR is comprised of refusals (including insufficient partials) divided by the total number of eligible units (including units with undetermined eligibility status). The major difference, of course, is that the IHSNG rates are based only upon outcomes the first time a unit is in sample.

Another important distinction is that the IHSNG rates make no attempt to standardize or define many concepts behind the outcome categories. For example, how does one classify a case where a household member is spotted inside the home by the interviewer, but he/she refuses to answer the door? We recognize that such situations will be classified as a refusal by some interviewers, as a no one home by others, and as a noninterview 'other reason' by still others. Similarly, housing units determined to be vacant, for example, may be classified for legitimate reasons as ineligible in one survey (and excluded from the nonresponse calculation) but eligible for other housing-unit based surveys where data are collected for vacant units (e.g., the American Community Survey and the American Housing Survey). The AAPOR definitions classify vacant units into the 'not eligible' category. The IHSNG did not make such a distinction and chose to accept such variations as legitimate survey design differences that must be considered (but not necessarily avoided) when interpreting and comparing initial contact rates. A detailed discussion of the initial contact response rates for the six surveys between 1990-1997 is documented in Atrostic et al., 1999.

Finally, we note that in addition to response rates and refusal rates, the AAPOR suggests two other measures not covered by the IHSNG core nonresponse set. These include cooperation rates and contact rates. Cooperation rates reflect the proportion of cases interviewed out of all eligible units ever contacted. AAPOR's Cooperation Rate 1 (COOP1)[2] can be approximated in an initial interview context using the following IHSNG variables and formula:

Initial Interview Cooperation Rate = NINT / NINT+NREF+NEUE+NEUL

Likewise, AAPOR's Contact Rate 1 (CON1)[3] measures the proportion of all cases in which some responsible member of the unit was reached and can be approximated in initial interview terms by:

Initial Interview Contact Rate = NINT+NREF+NEUE+NEUL / NEU

As illustrated here, although measures of cooperation and contact are not currently included in the IHSNG's recommended set of core nonresponse rates, they can easily be calculated using the set of variables outlined above.

4.2 Additional Recommended Outcome Information

As noted earlier, the core nonresponse rates are limited by the level of detail currently captured in the final noninterview outcome codes. A subgroup of the IHSNG recently completed a study of information on the reasons behind refusals, no one home, and other noninterview outcomes using informal and non-systematic information gathered during the course of surveys. These include interviewer-field communications (intercomms), interviewer notes, and previous coding studies. Having taken inventory of this information across several surveys, the group is now in the process of

---

[2]See AAPOR 1998, pg. 20.

[3]See AAPOR 1998, pg. 22.

recommending an expanded set of outcome codes for all automated surveys. Three new noninterview outcome codes being considered are:

     1) Health/mental problems (e.g., respondent too sick to respond, incapacitated);
     2) Access problems (e.g., gated community, buzzer entry, doorman, locked gate); and,
     3) Technical problems (e.g., computer problems).

Additionally, the group recommends recording case characteristics as a way to consistently gather and analyze reasons for nonresponse. According to this concept, a list of case characteristics will be displayed on a screen upon completion of an interview after selecting a noninterview classification code. Interviewers would record case characteristics by check boxes selectively displayed according to outcome code. For example, a sample of case characteristics available to check for refusals would include items such as: hostile respondent, respondent put off interview indefinitely, respondent too busy, refused because survey is voluntary, respondent has privacy concerns, questions too personal, etc.

In order to make the analysis of noninterviews more meaningful, the group has also recommended that similar case characteristics be recorded for *interview* outcome codes as well. The idea is to try and distinguish common characteristics of cases that eventually become noninterviews from those that eventually cooperate. Finally, the group is also working to develop a set of ancillary variables recorded for certain outcomes. Some examples include: number of contacts (recorded for interviews and non-interviews), signs of children being present, respondent demographics such as race, gender and age (for refusals), and whether or not the interviewer changed during the history of the current interview cycle (recorded for interviews and non-interviews). Assuming these recommendations are adopted, we will have good reason in the future to revisit the core set of initial interview rates and consider new ways to expand them.

5. Survey Specific Nonresponse Measures

A second IHSNG subgroup was charged with developing a set of nonresponse statistics appropriate to the goals of each survey. The task for this group was to determine a set of nonresponse measures specific to certain aspects of the survey design not reflected by the initial unit nonresponse rates. For example, in panel surveys, attrition over time is a concern due to the frequency of interviews conducted at the same household. Likewise, surveys that require self response from all household members need their own measure to benchmark the degree to which persons within a household are not interviewed (even though at the unit level, the household may be considered as an 'interview'). To this end, the IHSNG has thus far defined three survey specific nonresponse measures described below.

Variables Necessary to Compute Survey Specific Rates

Variable	**NEU**: Number of Eligible Units
Definition	Number of interviewing units in the sample that are considered to be eligible for interview, e.g., they are determined to be existing structures that are residentially occupied.

Computation	Sum of all units classified as eligible.
Details	Definition of 'eligible' may vary across surveys.

Variable	**NINT:** Number of Interviewed Units.
Definition	All interviews considered by a predetermined definition to be complete.
Computation	Sum of all interviewed units.

Variable	**NENIU:** Number of Eligible Non-Interviewed Households.
Definition	The sum of interviewing units eligible for interview that were not interviewed because of: language problems, no one home, temporarily absent, refusals, and all other reasons.
Computation	Sum of all eligible units classified as noninterviews

Variable	**NINP**: Number of Interviewed Persons.
Definition	All person interviews considered by a predetermined definitions to be complete in the current wave.
Computation	Sum of all interviewed persons

Variable	**NPNI:** Number of Persons Not Interviewed.
Definition	A person is classified as not interviewed if the unit provide at least one interview but another person within the unit is: 1) Never available, 2) A Refusal, 3) Physically or mentally unable with no proxy, or 4) Temporarily absent with no proxy.
Computation	Sum of all noninterviewed persons

Variable	**NEULU**: Number of Eligible Unable to Locate Units
Definition	Interviewing units eligible for interview which have moved to unknown addresses or moved more than 100 miles from the nearest primary sampling unit and a telephone interview cannot be conducted.
Computation	Sum of all eligible unable to locate units

Survey Specific Rates and Formulas

Rate	**PNR**: Person Nonresponse Rate
Definition	Number of persons not interviewed in interviewed units divided by the total number of eligible persons (interviews and noninterviews) in interviewed units.
Computation	NPNI / (NINP + NPNI) x100

Rate	**ULMR**: Unable to Locate Mover Rate
Definition	Number of unable to locate units divided by the total number of eligible housing units.
Computation	NEULU / (NINT + NENIU + NEULU) x 100

Rate	**SLR:** Sample Loss Rate
Definition	Number of eligible interviewing units that have attrited the sample up to and including the current wave (adjusted for growth) divided by the cumulative number of eligible interviewing units up to and including the current wave.
Computation	$\dfrac{\text{(NENIU + NEULU) * Growth factor}}{\text{(NENIU + NEULU) * Growth factor + NINT}}$
Details	In surveys that follow movers, more than one new address may result. When movers cannot be found, the numerator and denominator of the SLR formula are inflated with a growth factor to adjust for unaccounted addresses. This factor will vary from survey to survey.

For a presentation of trends in person nonresponse rates, unable to locate mover rates, and sample loss rates in the SIPP and NCVS during the 1990's, see Atrostic et al., 1999.

In addition to the three rates described above, we identified four more measures deemed useful to describe missing data and nonresponse in certain surveys. These included a person-level sample loss rate, proxy versus self-response rates, partial interview rates (sufficient and insufficient), and item nonresponse rates. To date, however, the interagency group has not yet completed the task of operationalizing the necessary concepts behind these measures or reached consensus regarding definitions and formulas.

A good example to illustrate the complexity of the task ahead is to consider the definition of an item nonresponse rate. On the surface, the measure seems fairly straightforward – to what degree did a survey fail to obtain an answer for a particular question? The reasons for item-level missing data are numerous – the interviewer could forget to ask the question or record the answer, the respondent could refuse to answer, the respondent could lack the information to answer, the interview could break off before all questions are asked, an automated instrument could have an error and not present a question in situation where it should, or questionnaire designers could choose to infer an item's value from other items rather than ask it directly.

In some cases (particularly those where the item is missing by design) the missing information can be logically inferred based on other information about the respondent or household. For example, if the marital status of the person listed as the spouse of the first person is missing due to nonresponse or instrument error, the entry can be logically edited from a blank to 'now married' based upon previous information. But the question arises, should the inferred answer be included or excluded when calculating an item nonresponse rate?

Data users may want to distinguish the number of cases where imputations are made from models or other formal imputation methods such as a hot deck procedures from the number of cases where the missing values can be reliably inferred based on other information collected in the interview. Typically the quality of the inferred information, particularly in the example noted above, is better than the quality of the fully imputed information. But questionnaire designers may think of item nonresponse

strictly as the percent of required responses that are missing for an item *prior to* any imputation or edits (i.e., based on uncoded and unedited datafiles). Of course the missing by design items confound this choice for measuring item missingness because they are handled by a logical assignment of values just like missing items can be. For example, in SIPP if a respondent says he/she owns an asset jointly with a spouse, the other married partner is not asked if he/she jointly owns also. This instrument design choice generates a missing data problem which is addressed through the logical assignment of information collected elsewhere in the instrument, just as we would infer the missing marital status information for the noncooperative spouse in the preceding example.

Aside from the example to compute "item missingness" discussed above, there is the issue of whether one reports cumulative measures of nonresponse, adding item nonresponse onto person nonresponse onto unit nonresponse. Since the "missingness" problem is handled very differently across these types of nonresponse and because the unit "missingness" solution is embedded within the weighting process, it's very difficult to produce the cumulative measure without having a misleading result when the sampling is not random or self-weighting.

Obviously, these examples of item nonresponse reflect very different measures each having a unique set of assumptions and definitions. The challenge is to establish common definitions and uses across surveys so the same concepts are applied and the resulting measure serves as an appropriate data quality indicator. Of course, even when we do arrive at agreement on how to measure item "missingness" consistently across surveys, we still may not have comparable measures. How do we compare degree of "missingness" associated with total income on a survey that has only one total income question to the degree of "missingness" for total income on a survey that measures over 50 different sources at the person level and then aggregates to compute total income? For example, if nonresponse in a detailed income survey is limited to a very minor income source (say a skipped $10 of interest income from an interest-bearing checking account), is that equivalent to skipping the answer for total income on a less detailed income survey? Alas, this is one of the items on our research agenda.

6. Getting Nonresponse Measures Routinely Produced

As mentioned previously, nonresponse rates in federal surveys are historically documented in reports such as the SIPP and American Housing Survey (AHS) quality profiles. But these publications are infrequent and not routinely produced for all federal surveys. Consequently, documentation of survey nonresponse measures are not currently aggregated into a common report format and are difficult to find. Moreover, the unit nonresponse measures published in these documents lack any type of standardization making comparison across surveys difficult and in some cases, misleading.

In the spring of 1999, the Census Bureau's Associate Director for Demographic Programs and Associate Director for Methodology and Standards called for development of a standard report on survey data quality. The agency formed an interdivisional committee to develop a standard profile that would accompany every microdata product the Census Bureau delivers to its clients and the public. This profile is to be referenced in every official report published from those demographic surveys.

The committee set about to create a user-oriented yet comprehensive profile designed to provide consistent and comparable quality measures across surveys and time. They produced an outline for a standardized quality profile that includes a section on data quality assessment. In addition to sampling error, coverage, and nonresponse adjustments, this section contains a subsection devoted to response rates. Specifically, the outline has a placeholder for reporting types and definitions of response rates, unit response rates and subunit rates (if applicable), attrition rates (if applicable), and item nonresponse rates. The quality profile package recognizes the IHSNG and its work to develop uniform methods for computing core nonresponse rates and directs managers to include these in this section of the profile.

The quality profile committee presented their initiative to Census Bureau managers and professional staff in mid-September, 2000. The suggested implementation plan calls for program areas to develop a system to begin routine preparation of the recommended data quality statistics in FY 2001, to develop standardized sections reporting on quality assurance measures in FY 2002, to review, revise and disseminate their first profile in FY 2003, and to begin routine production of the profiles in FY 2004.

Assuming the initiative is successful, we are hopeful that the core nonresponse measures developed by the IHSNG will soon be realized and become routinely available to better inform data users, survey managers, and survey sponsors of the status of nonresponse in federal household surveys. The development and application of survey-specific nonresponse statistics is still a work-in-progress. Where applicable, the surveys included in our study can begin to routinely produce person nonresponse rates, cumulative sample loss rates, and unable to locate mover rates – work remains to standardize additional measures beyond these.

References

American Association for Public Opinion Research. *Best Practices for Survey and Public Opinion Research and Survey Practices AAPOR Condemns*. Ann Arbor Michigan: AAPOR. 1997.

American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for RDD Telephone Surveys and In-Person Household Surveys.* Ann Arbor Michigan: AAPOR. 1998.

Atrostic, B.K, and Burt, G. *Household Nonresponse: What We Have Learned and a Framework for Further Work.* In Statistical Policy Working Paper 28: Seminar on Interagency Coordination and Cooperation. Office of Management and Budget: Washington, D.C. pp. 153-180. 1999.

Atrostic, B.K., Bates, N., Burt, G., Silberstein, A., and Winters, F. *Nonresponse in U.S. Government Household Surveys: Consistent Measures and New Insights.* A paper presented at the International Conference on Survey Nonresponse, Portland, Oregon, 1999.

Bailar, B. 'Is There a Unified Treatment of Response and Nonresponse Rates in Census Bureau Surveys?" Internal U.S. Census Bureau report, prepared by the Associate Director for Statistical Design, Methodology, and Standards, 1987.

Council of American Survey Organizations (CASRO), *On the Definitions of Response Rates*, Port Jefferson, N.Y., 1982.

Johnson, A.E., Botman, S.L., and Basiotis, P. *Nonresponse in Federal Demographic Surveys: 1981-199.* ASA proceedings of the Section for Survey Research Methods, pp. 983-988. American Statistical Association: Alexandria ,VA. 1994

Lynn, P., Laiho, J., Martin, J., and Beerten, R. *A Project to Standardise Response Rate Estimation in the UK.* A paper presented at the 11th International Workshop on Household Survey Nonresponse, Budapest, 2000.

Madow, W.G., Nisselson, H., Olkin, I., and Rubin, D.B., (editors), *Incomplete Data in Sample Surveys*, Vols. 1-3. New York: Academic Press, 1983.

Martin, M., and Straf, M.L. (editors), *Principles and Practices for a Federal Statistical Agency*. National Academy Press: Washington, D.C. 1992.

Office of Management and Budget. *Statistical Policy Working Paper on Measuring and Reporting the Quality of Survey Data (draft, June 2000).*

Office for National Statistics. *Statistical Quality Checklist.* Government Statistical Service: London, UK. 1997.

Shettle, C. Guenther, P., Kaspryzk, D., and Gonzalez, M.E. *Investigating Nonresponse in Federal Surveys*. ASA proceedings of the Section for Survey Research Methods, pp. 972-976. American Statistical Association: Alexandria ,VA. 1994.

U.S. Department of Commerce, Bureau of the Census. *SIPP Quality Profile 1998 - SIPP Working Paper No. 230.* Washington, DC. 1998.

U.S. Department of Labor, Bureau of Labor Statistics. *Current Population Survey Technical Paper 63 - Design and Methodology*. Washington, D.C. 2000.