HIGH RESPONSE ERROR AND POOR COVERAGE ARE SEVERELY HURTING THE VALUE OF HOUSEHOLD SURVEY DATA

Gary M. Shapiro and Donna Kostanich**
U.S. Bureau of the Census

I. INTRODUCTION

In this paper, we discuss two major sources of nonsampling error: undercoverage of the population and response error. We show that both of these can be quite substantial and result in major biases for some survey estimates. One of our central thesis is that these problems have received too little attention from survey sponsors and producers. Keane (1986) has made this same point, although he was addressing the entire nonsampling error field. Other than the decennial census, coverage has been rarely studied or even estimated. Response error has received more attention, and there have of course been numerous efforts to study and reduce particular response error sources. However, we believe there is not general awareness of deleterious effects of response error, and it is rarely estimated. In poorly designed and conducted household surveys, there can be many serious problems. In even the best household surveys, however, undercoverage and response error tend to be high and, in our opinion, are the two most important problems in the sample survey field.1/

We discuss the effects of undercoverage in section II of this paper, and of response error in section III. Section IV discusses important work that should be done on these errors. The concluding section summarizes how damaging these errors are and makes a plea for greater efforts to reduce them. Note that there is a more complete version of this paper available from the authors which, in particular, includes more details on the response variance model.

II. SURVEY COVERAGE

Although it is well known that household surveys and the census do not obtain complete coverage of the population, the magnitude of the coverage problem has been little publicized. We begin our discussion of coverage by describing how population figures are derived and used to estimate coverage. Section B discusses the causes and effects of undercoverage.

The reader will notice that we extensively quote Hainer et al (1988) in this section of the paper. The objectives and organization are completely different in the two papers, but we must cover much of the same material. Note also that the two papers have an author in common.

The Census Bureau regularly produces updated population estimates by age, sex, and race, and by ethnicity (Hispanic/non-Hispanic) by age and sex. The estimates are based on the most recent decennial census and updated each month to account for aging of the population, births, deaths, immigration, and emigration. They do not directly include an adjustment for census undercount. See Bureau of Labor Statistics (1987) for more details.

A. Magnitude of Undercoverage

The population estimates described above are used to routinely estimate coverage in Census Bureau surveys. This is done by calculating the ratio of the survey estimate of persons in a given age-sex-race group to the independent population estimate for that group. In producing estimates for the Current Population Survey (CPS), the Census Bureau gives a weight to each person which is inversely proportional to the probability of selection, multiplied by a noninterview adjustment. A ratio of 1.0 means that the survey coverage is the same as the census. A ratio of .8 means the survey coverage is 20 percent worse. Since noninterview rates are calculated, they do not contribute to coverage ratios being less than 1.0.

Average CPS coverage ratios by age, sex, and race/ethnic origin for 1986 are given in Tables 1 and 2. Coverage is about 7 percent worse overall in CPS than in the census. Table 1 shows that male coverage is worse than female coverage for all age groups for both whites and blacks.

Note that overall undercoverage for Black males is 17 percent worse than the census, and males 20-24 are 27 percent worse. Table 2 indicates that Hispanic undercoverage is apparently even worse than Black undercoverage.

It's important to note that the survey undercoverage shown in the Tables"... is in addition to decennial census undercoverage, which in 1980 was estimated to be about 1 percent overall and about 8.5 percent for Black males, according to demographic analysis (Fay, Passel, and Robinson, 1988). Thus, accounting for census undercoverage yields the following undercoverage rates for CPS: 8 percent for 14 and over, 25 percent for Black males 14 and over, at least 34 percent for Black males 20-24, and probably worse than 31 percent for Hispanic males." (Hainer et al, 1988)

Hainer et al (1988) show that CPS has better coverage than most Census Bureau surveys. For example, in 1978 coverage of nonwhite males is estimated to be 6 percentage points worse in the National Crime Survey than in the CPS.

Survey organizations other than the Census Bureau can estimate coverage just as the Census Bureau does, but rarely do. Some of the survey organizations who do calculate a ratio of survey estimates to independent adjusted census estimates do not previously multiply the survey estimates by a noninterview adjustment factor. Thus the calculated ratios show the combined results of noninterviews and undercoverage which make estimates of undercoverage impossible. We could find only two sets of non-Census Bureau data. Maklan and Waksberg (1987) state in general that "Most face-to-face surveys carried out by other organizations cannot achieve the CPS levels..." of coverage.

Cox and Cohen (1985) give data for the 1977 National Medical Care Expenditure Survey. Since this survey was partly conducted by National Opinion Research Center (NORC) and partly by Research Triangle Institute (RTI), two sets of coverage ratios are given. Each organization independently selected sample and collected data, so that differences in coverage ratios are not necessarily surprising. The coverage ratios in Tables 3 and 4 are derived from Table 4-2, p. 106-107, Cox and Cohen (1985). Looking first at coverage ratios for Whites, in Table 3, neither NORC nor RTI had any overall undercoverage for persons 15 and over for either males or females. Looking at age groups, however, indicates that both survey organizations tended to have overcoverage for older people (55+) and undercoverage for 15-54. Even for 15-54, however, undercoverage is less severe than in CPS.

For nonwhites, NORC and RTI achieved very different results. Table 4 indicates that NORC suffered undercoverage for nonwhites, comparable to CPS for females but much less severe for males. RTI, however, achieved substantial overcoverage for nonwhites, with no apparent undercoverage problem for any age group of either sex. Such overcoverage is so surprising, that it leads us to suspect that there may have been something in the field operations that led to overcoverage, possibly masking an underlying undercoverage problem.

In the only other data we have found, Maklan and Waksberg (1987) have concluded that random digit dialing surveys conducted by Westat generally have within household coverage at least as good as, and possibly better than, CPS. Their data shows, however, that Westat surveys also suffer from serious undercoverage.

A forthcoming paper (Cohen, 1988) will present coverage information for the National Medical Care Utilization and Expenditure Survey, a successor survey to the National Medical Care Expenditure Survey. Dr. Cohen has informed us that the paper will show undercoverage generally comparable to the levels of the CPS.

In summary, there are significant coverage deficiencies in Census Bureau household surveys and probably in most other household surveys as well. For Black males and for Hispanics, coverage is very poor.

B. Effects of Undercoverage on Data

1. Causes of Undercoverage

To understand the effects of undercoverage, it is useful to first discuss the causes. Undercoverage can occur because we miss entire housing units and because we miss some people within counted units. Not enough work has been done on coverage to quantify the relative severity of these two types of misses. We speculate, however, that entire unit misses are probably responsible for the majority of non-Black undercoverage, and that within unit misses are probably responsible for most Black undercoverage. We will briefly discuss the causes of both entire unit and within unit misses.

Entire unit misses occur fairly frequently during area listings. Most of the sample for Census Bureau surveys is selected directly from decennial census listings, but for rural areas and for most other organizations' surveys area sampling is used.

When sample is selected from census listings, there are also several ways units can be missed, the most obvious being when a unit was missed in the census. Sampling of census listings is supplemented in Census Bureau surveys by sampling from new construction permits, which also can result in missed units for several reasons.

We believe there are two main reasons for missing people within counted units. One of these is the deliberate omission of people. This can occur because presence of the missed people makes the family ineligible for the welfare benefits which they receive, because the missed people are engaged in illegal activities, because they are illegal aliens, or for other similar reasons. The second reason for missing people is the lack of fit between the "usual residence" concept used in many surveys and the actual living arrangements of some people. When people have connections with more than one address, there is the possibility of their being counted nowhere. See Hainer et al (1988) for detailed discussion of these reasons and the considerable qualitative evidence that they are important.

2. Effects Reported in the Literature

This section and the next discuss the effects on survey data resulting from missing people within counted units for the two reasons just discussed. We discuss some studies of such data effects, but there is little definitive information. Thus, the discussion in the next section is highly speculative.

The most obvious potential effect of undercoverage is on household composition data for the minority groups most seriously affected. In comparison with Census Bureau interviewers, Valentine and Valentine (1971) concluded that 12% of the sample households in their study were female-headed vs. a Census Bureau estimate of 72%. Though this was a small study and may be an extreme case, it suggests that surveys may substantially overstate the number of female-headed households, especially for Blacks.

Hainer et al (1988) state that the deliberate omission of men "... from the data is probably extremely biasing because the reasons they are missed are so directly related to important personal and household characteristics.... For instance, Clogg, Massagli, and Eliason (1986) discuss the implausible finding from the CPS that school enrollment rates are higher for Blacks than for Whites, for almost every age-residence category. They speculate that this occurs because of differential undercoverage of Black youth, with those attending school more likely to be counted than those who have dropped out." (Hainer et al, 1988)

There is a clearly implied link between deliberate omissions and sources of household income. One can expect "... that, in households depending on welfare, other sources of income will be under-or unreported..." (Hainer et al, 1988).

People who are missed because of no clear usual residence also probably result in significant bias. For example, "... Cook (1985) presents evidence suggesting that the National Crime Survey may underestimate the number of gun assaults by as much as one-third. He offers the explanation that the National Crime Survey does not adequately cover 'the kinds of people criminologists believe are most likely to be victims of serious violent crime--youthful males who are heavily involved in the life of the streets (including participation in criminal activity...)'" (Hainer et al, 1988)

In a study on the CPS, Hirschberg et al (1977) used a different estimation method than that normally used in the survey and compared results. Their method had two main differences from the standard method: (1) It adjusted for undercoverage in the 1970 Census rather than controlling to Census-levels unadjusted for census undercount; and (2) In the March supplement to CPS, special procedures are used to assure equality of husband's and wife's weights, in addition to controlling to age-sex-race figures as discussed above. The Hirschberg method was intended as an improvement over those procedures. The Hirschberg comparisons yielded substantial effects for aggregates, as would be expected since data were controlled to larger population figures. The effects on percentages and rates are much smaller, but in some cases significant; e.g., the unemployment rate increased from 4.5 percent to 5.0 percent and the poverty rate increased from 10.6 percent to 10.9 percent. Since Hirschberg, et al were attempting complex methodology, and since there can be no assurance that their methods are "correct" or "best," their results are difficult to interpret.

In an earlier CPS study, Johnson and Wetzel (1969) also found substantial effects on aggregates, but the unemployment rate was changed by only 0.1 percent.

3. Speculative Effects

In this section, we will give some examples of possible biasing effects on survey estimates. The assumptions in the second example are based on actual data, but the assumptions in the first example, though we believe them to be reasonable, are essentially unsubstantiated. The results given here should be regarded as illustrative of the biasing effects of coverage. We do not claim enough knowledge to present firm beliefs.

In the first example, the left-hand part of Table 5 is CPS supplement data on poverty rates for Black males. We believe that the poverty rate for missed Black males is much greater than for the covered population. We arbitrarily assumed rates consistent with our beliefs, as given in the second column. The resulting poverty rates for the combined covered and uncovered population are given in the last column. The overall poverty rate is increased by 5.1 percentage points, a 25% relative increase from the published data. For the widowed category, the increase is only 1.6 percentage points, while for "married, spouse present" the increase is 7.0 percentage points, a whopping 58% relative increase. In summary we speculate that undercoverage results in significant understatements of poverty rates for Black males, with some marital status categories much more affected than others.

In Table 6, speculated effects on family status are given. The data pertains to Black males 15 and over. Robert Fay has produced tables (unpublished) of undercoverage rates by various demographic groups. A match of April 1980 CPS with the Decennial Census was performed to identify persons in matched households who were interviewed in one source but not the other. In particular, Fay calculated net undercoverage rates for the CPS from this data for Black males 15 and over for four household relationship categories. These rates were used directly to compute the uncovered population percentages in the second column of Table 6. We used Fay's rates directly for all sub-categories that add to one of his four categories. In fact, of course, there are some sub-categories with worse coverage than the whole category, and this would make the uncovered population distribution more unlike the covered population than we assumed.

Fay's figures show a low rate of undercoverage for head/spouse and high rates for the other categories, resulting in some large differences for the combined population (column 3). For example, the percentage of Black males 15 and over who are householders drops from 37.4% to 31.7%. The percentage in the category :"Other unrelated persons in household" increases from 6.0% to 10.9%, a relative increase of 82%.

Also, the relationship between categories change. For instance, in the covered population there are many more nonfamily householders than other unrelated persons in households (13.6% vs.

6.0%), but for the combined population the estimates for these two categories are much closer (11.5% vs. 10.9%).

III. Response Errors

A. Background

In addition to the survey coverage problems discussed earlier, the precision of estimates is also affected by errors resulting from misclassification. The term "response error" is often used to describe these types of nonsampling errors. Response errors are not necessarily respondent error. They may also be a result of poor questionnaire design, faulty interpretation by the interviewer, or numerous other possibilities such as interviewing approach or attitudes. Even though the Census Bureau has historically been concerned with response error, there is still relatively little known about its effect on the data. For some of its surveys the Census Bureau does obtain estimates of response error; however, for most other survey organizations this is rarely done.

The Census Bureau utilizes an evaluation program known as "reinterview" in order to obtain estimates of both response bias and variance. The reinterview program was developed for the primary purpose of monitoring the interviewer's performance, so its design does not lend itself to estimating response error accurately. Theoretically in reinterview a different more experienced interviewer than the first interviewer conducts an identical but independent interview in a selected household.

For all practical purposes, the reinterview cannot be conducted under identical conditions, nor can responses truly be independent since respondents often recall their earlier answers. Although not satisfied, these conditions are nonetheless important in obtaining good estimates of response error. See U.S. Bureau of the Census (1978) for more details about the reinterview program. In spite of these limitations and the typically small sample that is reinterviewed, the reinterview program has provided useful information on the magnitude of both response bias and variance.

At the Census Bureau, reinterviews are conducted on a regular basis for most major surveys such as the Current Population Survey (CPS), the American Housing Survey (AHS), the National Health Interview Survey (NHIS) and the Survey of Income and Program Participation (SIPP). Only the CPS reinterviews consist of a complete coverage of questions. For SIPP, a very limited reinterview is done for only a small number of questions which are asked in a slightly different manner. We suspect that due to lack of independence, reinterviewed results probably understate the actual levels of response bias in the surveys.

On the other hand, estimates of response variance, as measured by reinterview, are often quite high (i.e. accounting for 20 to 50% of the total variance). For CPS, response variance has been estimated to account for nearly a third of the total variance for unemployed. See Newbrough (1988). For AHS questions relating to housing unit conditions or neighborhood services, response variance almost always accounts for 40, 50 or 60 percent of the total variance. See Schwanz (1986). For example, about 60 percent of the variance on whether there are satisfactory hospitals or health clinics nearby is response error.

We are not very concerned with the effect of these high levels of response variance on simple totals and proportions. At least for simple random sampling, McCarthy (1969) shows that response variance does not increase the total variance of simple totals and proportions. We are, however, very concerned with what impact this has on the expected value of more complex estimates, such as crosstabulations, measures of association between variables, regression, and log-linear analysis. Under some circumstances high response variance could lead to serious biases. Fuller (1986) illustrates how response error affects gross change in employment status. Other work on response error models for survey data has also been conducted. See Dalenius (1977) for a bibliography of work in this area.

B. Simplified Model

This section examines the potential impact only of response variance. Although we are very concerned about response bias, the affects of response variance are less well known. Here we consider a simple but somewhat realistic situation to examine the impact of response variance on biases for 2 x 2 crosstabulations. X and Y will represent the two characteristics of interest, each with two categories (i.e. 0 or 1). Their simple crosstabulation, for the full population with no response error, is represented by:

Y \ X	1	0	
1	Z(1,1)	Z(1,0)	Y(1)
0	Z(0,1)	Z(0,0)	Y(0)
	X(1)	X(0)	N

In order to examine biases associated with this table we make the following assumptions: 1) There is response error associated with variable X but <u>not</u> for variable Y; 2) The expected response for X is P(i) if X = i; where P(i) is between 0 and 1; and 3) There is no overall response bias, i.e., the net effect of response error is zero.

Using these assumptions and further assuming that the whole popualtion is interviewed we looked at the relative biases associated with the above crosstabulation. The maximum of the absolute value of the resulting relative biases for the estimated cell frequencies, $\hat{Z}(i,j)$ turns out to be a function of only the following parameters:

1) P(i,1) = Z(i,1)/Y(i) for i = 0 or 1
2) P(x) = X(1)/n

and 3) Q(1) = 1-P(1) (this gives the rate of misclassification for X = 1 population values).

We determined maximum absolute relative biases, for various values of the above 3 parameters, the results of which are given in Table 7.

From this table we first note that when P(i,1) = P(x) the relbias is zero for all values of Q(1). As the difference between P(x) and P(i,1) increases the relbiases also increase in absolute value. This seems to be saying that response variance will have a more negative impact on more highly correlated data and this impact is greater for larger values of Q(1). Note that when the true correlation is zero, P(1,1) = P(0,1) and when this occurs P(1,1) = P(0,1) = P(x).
As the difference between P(1,1) and P(0,1) increases, so does the difference between P(1,1) and P(x). This has the effect of increasing the magnitude of the population correlation.
Consequently, variables that are more highly correlated will be affected more by response error.

Note when Q(1) is larger than .2 (i.e. more than a 20% chance of a misclassification error), nearly all the biases are quite serious. Note that even when the misclassification error rate is only 10% (i.e. Q(1) = .1) there are still many situations that can result in serious biases. For example, when P(i,1) = .8 and P(x) = .3 the maximum absolute relative bias is .36. Thus, an estimated cell frequency would be off by 36%.

We had also looked at biases for smaller misclassification errors. Relative biases were simulated as discussed above but we used values of Q(1)<.1. These simulated results are given in the larger version of this paper. Similar types of relationships as seen in Table 7 were found but the magnitude of the relative biases were not nearly as large. Many of these relative biases were less than .2, but some were still quite severe even for very small Q(1). For example, when the misclassification error is only 3% and when P(i,1)=.9 and P(X)=.3, the relative bias is .26 in absolute value.

C. Effect on Data - Examples

This section examines how biased published estimates could be if the model discussed in the previous section were valid. To do this we let the published estimates represent the estimated cell frequencies, $\hat{Z}(i,j)$'s, discussed earlier. Using actual reinterview results we can estimate Q(1), the misclassification error, and then derive what the true $\hat{Z}(i,j)$'s would be under this particular model.

From reinterview data we estimate the proportion of the total variance that is response variance. This is obtained from the index of inconsistency measure I. Given I, we get $Q(1) = Q(x)[1- \sqrt{1-I}]$

Example 1

From CPS published estimates found in U.S. Bureau of Labor Statistics (1988), we formed the following crosstabulation of unemployed/employed persons by two major occupation categories: 1) Managerial and Professional (Man/Prof) and 2) Operator, Fabricators and Laborers (Oper/Lab).

Published Estimates (in thousands)

	Unemployed	Employed
Man/Prof	615	28503
Oper/Lab	1998	17207

To obtain the true populations values under the model, the reinterview estimate of the index of inconsistency for unemployed was used. See Newbrough (1988). This gives I=.33, which results in a misclassification error of Q(1) = .17. These values and the published information above give the following true cell values under the assumed model:

True Values (in thousands)

	Unemployed	Employed	
Man/Prof	402 (.53)	28716 (-.01)	r= -.81
Oper/Lab	2211 (-.10)	16994 (.01)	

The numbers in parentheses give the relative biases for the published estimates. Consequently if the model were correct, the published estimate of 615,000 would overestimate the true value of unemployed manager and professionals by 53%. Although the other cells are affected by a small degree, this cell reduces the magnitude of the correlation from its real value of -.81 to -.69.

Example 2

Next we examine estimates from the AHS, published in U.S. Department of Commerce and U.S. Department of Housing and Urban Development (1984). The following shows the relationship between housing units with broken plaster and whether these housing units are in metropolitan areas.

Published Estimates (in thousands)

	Broken Plaster	No Broken Plaster	
Met	634	5528	r = .26
NonMet	66	973	

From reinterview we estimate that about 50% of the total variance for broken plaster is due to response error, thus I = .5. This yields a misclassification error Q(1), of .26. Again under the assumed model, this information yields the following true population values.

True Values (in thousands)

	Broken Plaster	No Broken Plaster	
Met	648 (.02)	5514 (.003)	r = .39
NonMet	52 (.28)	987 (-.01)	

As before, the relative biases are included in parentheses. Once again, one cell is considerably more biased than the others. This example shows that we could be overestimating housing units with broken plaster in nonmetropolitan areas by 28%. As in the first example, this example also shows that the published data may be understating how strongly related the two variables really are. The correlation for the published estimates was computed to be only .26 while the true correlation is .39.

IV. RESEARCH TO IMPROVE COVERAGE AND REDUCE RESPONSE ERROR

We have shown that undercoverage and response error cause serious biases, and that we lack understanding of the causes of the problems. In this section, we briefly discuss some work that should be done to improve our knowledge.

A. Coverage Work

Hainer et al (1988) discussed a four part research program. See that paper for details. The four parts of the program are expanded participant observation research, the use of existing CPS and Census data to study coverage, conducting debriefing studies and ethnographic research to learn about household structure, and experimenting with several innovations intended to improve interviewer performance and confidentiality perceptions. We believe that this research can not only improve our understanding of undercoverage but also significantly reduce it.

B. Response Error Work

Before we discuss ways to reduce response error, we must first emphasize the importance of obtaining good estimates of response error components. As noted earlier, most survey organizations do not even attempt to measure this type of error. Although the Census Bureau does estimate these errors for some of its surveys, the results provide only limited information. Until we can measure response error reasonably well, we can not thoroughly understand how it could be distorting the data let alone determine how to reduce its effect. The following briefly outlines some steps that could be taken to improve our understanding of response error.

1. Improve Reinterview

The first step would be to expand or establish a reinterview program that is specifically geared towards producing good estimates of response error. The reinterview sample should be designed to increase the precision of response error measurements. To estimate response variance we would want to exert more control over reinterview in order to minimize, to the extent possible, any procedural differences between the original interview and the reinterview. To estimate response bias we would want the reinterview to produce the most accurate responses possible, even if this means a change in procedure from the original interview. Additional knowledge about response error could also be obtained by conducting a detailed analysis of the reinterview data, such as O'Muircheartaigh (1986) did for CPS. This study compared response error measurements by demographic characteristics and interview procedures such as proxy vs. self response.

2. Conduct Cognitive Research

Cognitive research is an area we think may provide a lot of insight into the causes of response error. This type of research explores the cognitive processes of respondents during the interview and could be used to identify where and why measurement errors occur. We could obtain information on which questions are ambiguous or conceptually difficult to interpret or require an unusual amount of recall. Questions could then be revised and iteratively retested using these cognitive processes until reasonably precise questions are arrived at.

3. Reduce Response Variance

Procedural changes other than improving the questionnaire could also be investigated to reduce response variance. Proposed procedural changes could be examined using carefully controlled reinterview experiments. Some suggestions which have recently been made for SIPP (Singh (1988)) include the following:

- provide interviewers with more cues and probes.
- use different procedures for different types of respondents (e.g. more probes, more follow-up.)
- more emphasis on data quality.
- provide interviewers with advance feedback.
- perform an extensive edit and followup in the field.
- obtain a statement of commitment from respondents.

V. CONCLUSION

We have separately discussed the magnitudes and effects of 2 major types of error, undercoverage and response variance. We have shown that we miss over 30% of the persons in some demographic groups, and that as much as 60% of the total variance is due to response variance for some characteristics. We have discussed probable major biasing affects on household composition, crime victimizations, and poverty rates due to undercoverage. We have also discussed probable major biasing affects on cross-tabulations of labor force category by occupation and of housing conditions by geography due to response variance. There is also a combined effect of undercoverage and response variance which makes matters worse. The very demographic groups that have high rates of undercoverage are probably also subject to particularly high response variance.

For example, for Black males 20-24, we may frequently get proxy responses from people who are uncertain of their activities, and for Hispanics, language problems likely lead to relatively high response errors. Thus, the combined biasing effects of response error and of undercoverage can be much worse than the separate effects.

We believe it is vital to make much more effort to understand and reduce these nonsampling errors. We emphatically agree with Keane (1986) who stated: "Insufficient funding is a ... factually correct barrier to increasing knowledge about nonsampling error measurement." We have briefly discussed many things that could be done. It at least ought to become standard practice to measure the magnitudes of these two errors. They are rarely measured outside the Census Bureau, and inconsistently within.

With some substantial resources devoted to undercoverage and response error, we believe they could be significantly reduced, thereby resulting in major improvements in data quality. We hope survey organizations and survey sponsors will make needed resources available.

VI. Acknowledgement

We thank Paul Bettin, Robert Fay, M. James Roebuck, and Robert Smith, all of the Census Bureau, for their useful comments. All errors and opinions, however, are solely the authors. We would also like to thank Jamila Woods and Cora Wisniewski for their typing of the paper.

REFERENCES

[1] Bailey, R., Schreiner, I., Baer, L., and Kerstetter, L. (1968), "The Current Population Survey Reinterview Program January 1961 through December 1966," Technical Paper 19, U.S. Bureau of the Census, December 1968.

[2] Clogg, C., Massagli, M., and Eliason, S. (1986), "Population Undercount as an Issue in Social Research," Proceedings of the Second Annual Census Bureau Research Conference, U.S. Bureau of the Census, Washington, D.C.

[3] Cohen, S. (1988), "Household Effects in the National Medical Care Utilizaition and Expenditure Survey," paper to be presented at the American Public Health Association Annual Meeting.

[4] Cook, P. (1985), "The Case of the Missing Victims: Gunshot Woundings in the National Crime Survey," J. Quantative Criminology, 1:19-102.

[5] Cox, B.S. and Cohen, S.B. (1985), Methodological Issues for Health Care Surveys, New York: Marcel Dekker Inc.

[6] Dalenius, T. (1977), "Bibliography of Nonsampling Errors in Surveys," International Statistical Review 45, 71-81 and 181-197.

[7] Fay, R. (1988) "A Preliminary Analysis of Household Undercoverage in the 1980 Current Population Survey," draft internal Census Bureau memorandum to G. Shapiro and T. Moore, July 6, 1988.

[8] Fuller, W.and Chua, T. (1986) "Response Error Models for Changes in Multinominal Variables," Proceedings of the Second Annual Research Conference, U.S. Bureau of the Census, Washingotn, D.C. pp. 425-441.

[9] Hainer, P. (1987), "A Brief and Qualitative Anthropological Study Exploring the Reasons for Census Coverage Error Among low Income Black Households," report submitted to the U.S. Census Bureau, April 1987.

[10] Hainer, P., Hines, C., Martin, E., and Shapiro, G., (1988), "Research on Improving Coverage in Household Surveys", Proceedings of the Census Bureau Fourth Annual Research Conference, p. 513-539.

[11] Harwood, A. (1970) "Participant Observation and Census Data in Urban Research," paper presented at the annual meeting of the American Anthropological Ass'n, San Diego, California, November 1970.

[12] Hirschberg, D., Yuskavage, R. and Scheuren, F.J., (1970), "The Effect on Personal and Family Income of Adjusting the CPS for Undercoverage," Proceedings of the Social Statistics Section of the American Statistical Ass'n Meeting, pp. 70-80.

[13] Johnston, P.F. and Wetzel, J. (1969), "Effect of the Census Undercount on Labor Force Estimates" Special Labor Force Report No. 105, Bureau of Labor Statistics.

[14] Keane, J. G. (1986) "Have We Erred in Nonsampling Error?" Proceedings of the Census Bureau Second Annual Research Conference, p. 183-189.

[15] Kniesner, T.J., McEroy, M.B. and Wilcox, S.P. (1988), "Family Structure, Race and the Hazards of Young Women in Poverty," paper presented at the Conference on Individuals and Families in Transition: Understanding Change Through Longitudinal Dat, March 1988.

[16] Maklan, D. and Waksberg, J. (1988) "Within Household Coverage in RDD Surveys" to be published in Telephone Survey Methodology, John Wiley and Sons.

[17] McCarthy, P.J. (1969), "The Effects of Discarding Inliers When Binomial Data are Subject to Classification Errors," Research Triangle Institute Technical Report No. 3 prepared for U.S. Bureau of the Census, October 1969.

[18] Newbrough, J. (1988), "Report of CPS Reinterview for the First Half of 1987," Unpublished U.S. Census Bureau memorandum.

[19] O'Muircheartaigh, C. (1986), "Correlates of Reinterview Response Inconsistency in the Current Population Survey,"Proceedings of the Second Annual Research Conference, U.S. Census Bureau.

[20] Schwanz, D. (1986), "Minutes of the August 14, 1986 Meeting of the AHS Response Variance/Bias Research Work Group," unpublished U.S. Bureau of the Census memorandum.

[21] Singh, R.P. (1988), "Minutes of June 7, 1988 Meeting", unpublished Census Bureau memo.

[22] U.S. Bureau of the Census (1986) "Marital Status and Living Arrangements: March 1985," Series P-20, #410, November 1986.

[23] U.S. Bureau of the Census (1978), The Current Population Survey: Design and Methodology, Technical Paper No. 40, Washington, D.C.

[24] U.S. Bureau of the Census (1987), "Poverty in the United States 1985," Series P-60, #158, October 1987.

[25] U.S. Bureau of Labor Statistics (1988), Employment and Earnings, Vol. 35 No. 1.

[26] U.S. Bureau of Labor Statistics (1987), Employment and Earnings, Vol. 34 No. 12.

[27] U.S. Department of Commerce and U.S. Department of Housing and Urban Development (1984), Current Housing Reports, Series H-150-83, General Housing Characteristics.

[28] Valentine, C. and Valentine, B.L. (1971) "Missing Men: A Comparative Methodological Study of Underenumeration and Related Problems," unpublished report submitted to the U.S. Census Bureau, May 3, 1971.

** This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author(s) and do not necessarily reflect those of the Census Bureau.

1/In some surveys, small sample size is a more serious problem but the resulting high sampling errors are well-understood and documented.

TABLE 1: COVERAGE RATIOS BY AGE FOR CURRENT POP. SURVEY AVERAGE FOR 1986

	Total 14+	14-19	20-24	25-44	45-64	65+
Total	.932	.946	.887	.924	.935	.967
White						
Male	.925	.950	.885	.914	.936	.946
Female	.950	.951	.919	.946	.946	.986
Black						
Total	.874	.904	.778	.856	.884	.946
Male	.833	.884	.733	.805	.861	.927
Female	.907	.924	.820	.910	.906	.956

TABLE 2: COVERAGE RATIOS FOR HISPANICS BY AGE AND SEX FOR CPS AVERAGE FOR 1986

	Total 14+	14-19	20-29	30-49	50+
Total	.798	.845	.769	.808	.800
Male	.773	.870	.731	.762	.782
Female	.823	.820	.792	.853	.816

TABLE 3: COVERAGE RATIOS BY AGE FOR WHITE PERSONS, NATIONAL MEDICAL CARE EXPENDITURE SURVEY

	Total 15+	15-54	15-24	25-44	45-54	55-64	65+
Males, NORC	1.001	.965	.995	.971	.937	1.117	1.096
Males, RTI	1.015	.982	1.065	.948	.954	1.074	1.158
Females, NORC	1.011	.972	1.036	.943	1.012	1.083	1.122
Females, RTI	.997	.964	1.065	.957	.986	1.001	1.139

NOTE: Coverage Ratios computed from data in Table 4-2, p.106-107, Cox and Cohen (1985)

TABLE 4: COVERAGE RATIOS BY AGE FOR NONWHITE PERSONS, NATIONAL MEDICAL CARE EXPENDITURE

	Total 15+	15-24	25-44	45-64	65+	0-14
Males, NORC	.936	1.036	.921	.898	.771	.971
Females, NORC	.944	.990	.925	.966	.843	.951
Males, RTI	1.129	1.065	1.207	1.014	1.326	1.116
Females, RTI	1.120	1.128	1.131	1.072	1.157	1.140

NOTE: Coverage Ratios computed from data in Table 4-2, p. 106-107, Cox and Cohen (1985)

Table 5 - % Below Poverty Status, By Marital Status, in 1985 of Black Males 15 and Over

	Covered Population 1/	Noncovered Population	Combined, Covered & Noncovered
Total, 15 Yrs. & Over	20.2	-	25.3
Single	26.5	45	31.1
Married, Spouse Present	12.0	40	19.0
Married, Spouse Absent	25.1	45	30.1
Separated	25.1	45	30.1
Other	24.8	45	29.8
Widowed	28.4	35	30.0
Divorced	20.0	35	23.8

1/ Source of Data: Table 8 of U.S. Census Bureau (1987)

Table 6 - Family Status of Black Males, 15 and over, for Covered, Uncovered, and Total Population: March 1985 (%Distribution)

	Covered Population 1/	Uncovered Population	Combined Covered and Uncovered
Black Males	100.0	100.0	100.0
In Families	79.7	66.1	76.3
Householder	37.4	14.6	31.7
Married, Wife present	33.7	13.2	28.6
Other	3.8	1.5	3.2
Husband of householder	4.4	1.7	3.7
In related subfamily	1.6	4.6	2.4
Married, wife present	0.7	2.0	1.0
Parent, no wife present	0.3	0.9	0.4
Child	0.6	1.7	0.9
Child of Householder, not in related subfamily	29.5	25.9	28.6
Other, not in related subfamily	6.7	19.3	9.8
In unrelated subfamilies	0.3	1.3	0.5
Married, wife present	0.1	0.4	0.2
Other Reference person	0.1	0.4	0.2
Child	-	-	
Other, in unrelated subfamily	0.2	0.9	0.4
Not in Family Group	20.0	32.5	23.1
Nonfamily Householder	13.6	5.3	11.5
Other Unrelated persons in Household	6.0	25.5	10.9
In group quarters	0.4	1.7	0.7

1/ Source of data: Table 2, p. 25 of U.S. Census Bureau (1986)

Table 7 Maximum Absolute Value (Relbias Z(i,1), Z(i,o))

P(i,1)	P(x) \ Q(1)	0.1	0.2	0.4	0.6	0.8	1
0.1	0.1	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.3	0.29	0.57	1.14	1.71	2.29	2.86
0.1	0.5	0.80	1.60	3.20	4.80	6.40	8.00
0.2	0.2	0.00	0.00	0.00	0.00	0.00	0.00
0.2	0.4	0.17	0.33	0.67	1.00	1.33	1.67
0.3	0.1	0.07	0.15	0.30	0.44	0.59	0.74
0.3	0.3	0.00	0.00	0.00	0.00	0.00	0.00
0.3	0.5	0.13	0.27	0.53	0.80	1.07	1.33
0.4	0.2	0.06	0.13	0.25	0.38	0.50	0.63
0.4	0.4	0.00	0.00	0.00	0.00	0.00	0.00
0.5	0.1	0.09	0.18	0.36	0.53	0.71	0.89
0.5	0.3	0.06	0.11	0.23	0.34	0.46	0.57
0.5	0.5	0.00	0.00	0.00	0.00	0.00	0.00
0.6	0.2	0.12	0.25	0.50	0.75	1.00	1.25
0.6	0.4	0.08	0.17	0.33	0.50	0.67	0.83
0.7	0.1	0.22	0.44	0.89	1.33	1.78	2.22
0.7	0.3	0.19	0.38	0.76	1.14	1.52	1.90
0.7	0.5	0.13	0.27	0.53	0.80	1.07	1.33
0.8	0.1	0.39	0.78	1.56	2.33	3.11	3.89
0.8	0.3	0.36	0.71	1.43	2.14	2.86	3.57
0.8	0.5	0.30	0.60	1.20	1.80	2.40	3.00
0.9	0.2	0.88	1.75	3.50	5.25	7.00	8.75
0.9	0.4	0.83	1.67	3.33	5.00	6.67	8.33