CURRENT POPULATION SURVEY REPORTING OF SOCIAL SECURITY NUMBERS

Linda Vogel and Terry Coble
Social Security Administration

The Social Security Administration and the Census Bureau are engaged in a project in which SSA earnings and benefit records are being matched to Current Population Survey data. (See the introductory comments to this session.) The goal of this effort is an improved data base for use in studies on income distribution and redistribution. In order to locate SSA earnings and benefit data for an individual, one must have, or be able to determine, his social security number (SSN). Of primary importance in the 1973 Match Project, therefore, was the extent of missing SSN's in the Current Population Survey.

This paper examines the reporting and nonreporting of SSN's from various vantage points. In section 1, a description is given of how the SSN is collected in the CPS and the possible sources of error that can be introduced. Section 2 discusses the overall response rate to the SSN question and compares our results to previous studies. The remaining sections analyze some of the characteristics of the individuals in the survey which may influence the likelihood of an SSN being reported (section 3); examine the characteristics of the household respondents which affect whether a number is reported (section 4); and present some other aspects of SSN reporting, such as differences by data collection centers and by rotation groups (section 5).

1. COLLECTION OF SOCIAL SECURITY NUMBERS IN THE CURRENT POPULATION SURVEY

Background.--The structure of the CPS since the early 1950's has been such that each sample household is eligible for interview on eight separate occasions: four consecutive months one year, followed by the same four calendar months a year later, with an eight month "rest" in between. These interviews are staggered so that for any given month's survey--March 1973, for instance--one-eighth of the cases are being enumerated for the first time, one-eighth for the second, and so on. Thus, each month's sample actually consists of eight subsamples, or "rotation groups," which can be distinguished by the number of times the addresses they contain have been eligible for interview. 1/

A "control card" is prepared for a household when it first enters the survey. The control card is a basic repository of data and lists for each household member such items as name, date of birth, age, race, sex, etc. In 1963, in connection with the first SSA-CPS linkage project [102, 118, 119, 121, 147, 148], the control card was revised to obtain social security numbers for persons 14 years or older. Since then, SSN information has been routinely requested just of rotation groups which enter the sample during the months of December, January, February, or March. The reason for this is that interest in linking SSA and CPS data, as mentioned in the overall introduction to this session, has centered on comparing income responses in the CPS with the corresponding SSA administrative figures. 2/ Rotation groups entering in these four months are the only ones in the survey in March when the supplementary income questions are asked.

The exact phrasing of the SSN question on the control card is, "What is . . . .'s Social Security (Railroad Retirement) number?" Interviewers are instructed to tell respondents, if asked, that the SSN's are necessary to help evaluate the accuracy, consistency, and comparability of some of the statistics produced from the CPS; that this evaluation is conducted using various techniques, one of which is through the use of governmental administrative records; and that these evaluation programs are carried out by sworn employees of the Census Bureau under the strictest controls, to protect the confidentiality of the information secured. As an example, respondents are told about the possibility of matching the survey data to the records of the Social Security Administration.

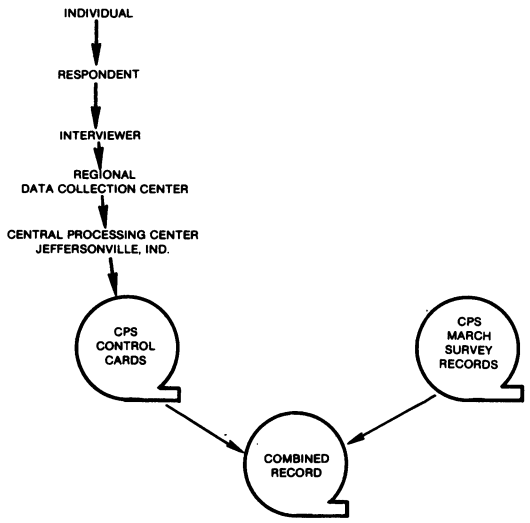
Usually, in the CPS, one member of a household responds for the entire household. If the household respondent knows, or thinks he knows, the SSN's for the other members, these are recorded by the interviewer on the initial visit. If the respondent reports that a person has no number, "none" is recorded. The respondent may, at this time, of course, refuse to provide numbers. 3/ When the respondent is unaware of the SSN's of other household members, a form is left to be filled in and mailed to the census offices. If any numbers are still missing on the second or later months of interview, they are again requested.

SSN Collection Process.--The major steps in the collection of social security numbers are set forth in figure 1. At each of these steps, errors of response or nonresponse can be introduced which would affect the quality and completeness of the SSN reporting. The individual, himself, is obviously a possible source of error. He may not know or may misremember his SSN. Another source of error can arise if the individual does not respond for himself, but the information is obtained by proxy. This was the case for

approximately 57 percent of the individuals in the survey who did not answer the SSN question directly--it was answered for them by someone else in the household. Despite careful training and control, the CPS interviewers may also have introduced some errors, either by failing to ask the question at all or by transcription mistakes and the like. Differences among the Census Bureau's regional data collection centers in the emphasis placed on the SSN question are, also, a possible cause for concern.

Up to this point, the steps, as shown in figure 1, to obtain SSN data in the CPS do

Figure 1.—PROCESSING FLOW OF SSN'S



not differ very much from the steps that would be taken for any other item included in the survey. Unlike much of the other information on the control cards, however, SSN's are not routinely transcribed onto the monthly survey questionnaires. For this project, therefore, the control cards had to be specially encoded into machine-readable form (at the Bureau's Central Processing Center in Jeffersonville, Indiana). They were subsequently matched to the tape files of the March 1973 interview schedules in order to have both the CPS income data and the SSN information available for an individual on one combined record. These latter steps are, of course, also subject to clerical and other errors. For example, lost control cards would result in missing SSN's.

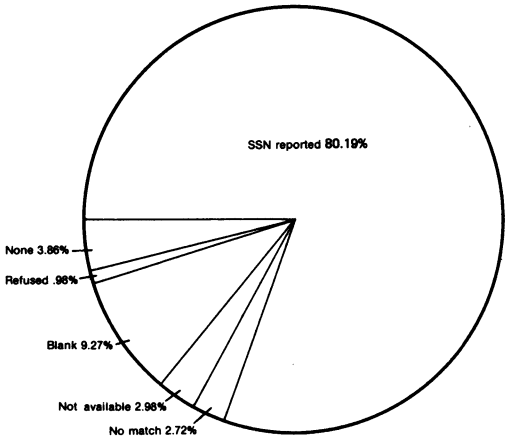
In this paper, as was previously stated, we will be confining our attention to comparisons between persons who have had an SSN recorded for them in the CPS and those who have not. The question of the correctness of the SSN responses given is not taken up here. (The steps taken to validate numbers and to search for missing and misreported SSN's are covered in the subsequent papers at this session.)

For our purposes, all numerical responses to the SSN question, whether valid or invalid, have been treated as equivalent. Since essentially the entire U. S. population 17 years or older has an SSN, 4/ instances in which no SSN was reported in the CPS for such persons are, for the most part, cases in which an error of some sort was made. While we cannot always relate our hypotheses regarding the completeness of the SSN responses to specific steps listed in figure 1, nonetheless, the reader may find it useful to keep the various steps in mind during the discussion which follows.

2. OVERALL RESULTS AND COMPARISONS TO OTHER STUDIES

Four out of five of the over 100,000 CPS persons 14 years or older in households interviewed in March 1973 reported a social security number (see figure 2). Of

Figure 2.—OVERALL RESPONSE TO SSN QUESTION



the remaining 20 percent who were not shown to have reported SSN's, the response "none" was recorded for 3.9 percent, while only one percent were listed as refusals. Blank responses, possibly due to the interviewer not emphasizing or asking the SSN question at all, totaled approximately 9.3 percent. Answers of "not available" are recorded for almost three percent of the cases. This situation might arise if the respondent failed to mail in the form left by the interviewer for numbers not obtained at the time of interview or if SSN's on the returned forms were never incorporated on the control card. Finally, missing numbers resulted for about 2.7

percent of the March supplement records because no match could be made to the source of the SSN, the control card.

The percentage of social security numbers reported in the CPS, about 80.2 percent, represents a successful effort on the part of the Census Bureau in collecting the numbers and is an improvement over previous comparable projects. In the 1963 Pilot Link Study, 72 percent of the 15,533 persons 14 years or older initially responded with an SSN [147]. A similar 1966-67 Census Bureau study by Ono, Patterson, and Weitzman [93] found that approximately 76 percent of the 4,500 persons 14 years or older comprising the sample reported an SSN.

While it is enlightening to compare the positive response rate of the Match project with previous studies, it is necessary to remember that the present study should yield an increase in the percentage of SSN's reported for two reasons. First, the SSN is more widely used now than ever before, for such diverse purposes as automobile driver's licenses and student identification, etc. One would expect, therefore, that more people are able to report their social security number than in the past. Secondly, a larger proportion of the U. S. population now has SSN's. This increase is especially noticeable among women, as this segment of the population has joined the labor force in greater numbers. Although for the above-mentioned reasons we would expect a higher SSN reporting rate in the current study, there is, on the other hand, another factor involved that should be taken into account; namely, the growing tendency of individuals to be reluctant to divulge what they consider to be confidential information and to regard such questions as an invasion of privacy. Fortunately, judging from the low refusal rate, this still does not seem to be a significant factor in Census Bureau conducted surveys involving the social security number.

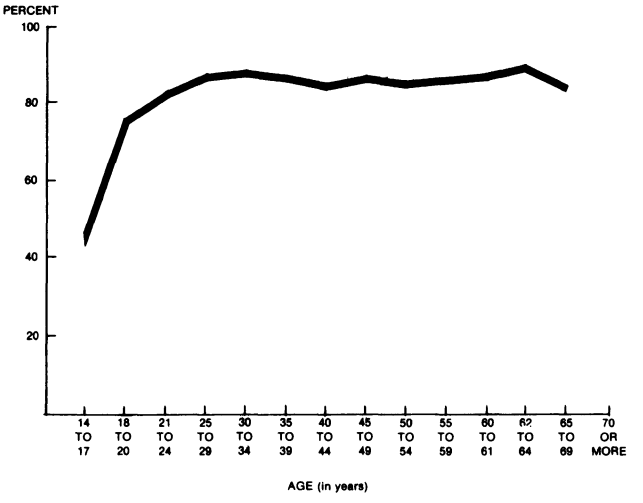
3. SSN REPORTING BY AGE, RACE, SEX, AND RESPONDENT STATUS

Individual's Age.--Age is found to be, by far, the most significant characteristic in determining whether an SSN is reported for an individual. 5/ Persons 14 to 20 years old appear to be much less likely to show a positive SSN response; this was especially true of those in the 14 to 17 year-old category, where only 46.0 percent had an SSN reported (see figure 3).

Age also has a strong influence on whether or not an individual shows up in the "none" or "blank" groups, but it seems to have little impact on the proportion of "no matches" or answers of "not available." Approximately 23.0 percent of all 14 to 17 year-olds were shown to have an answer of "none" for the SSN question, compared to 1.5 percent for the other age groups. The age effect on the number of blank entries is also striking; for persons younger than 21 years of age, almost 20 percent were found to have blanks, while the percentage drops to seven percent for persons 21 or older.

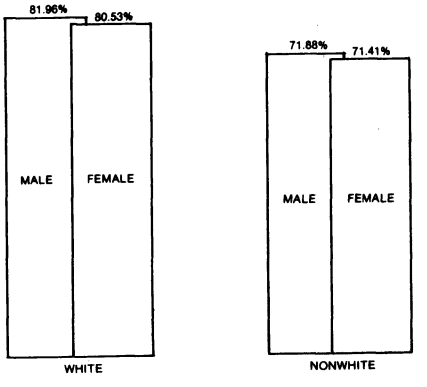
Refusals, though a small proportion of the total sample, behave in an interesting way. Those over the age of 35 are more likely to be listed as refusals (1.3 percent) than are persons in the younger age groups (0.7 percent).

Figure 3.—SSN REPORTING BY AGE



Individual's Race and Sex.--Proportionately more whites than nonwhites have an SSN reported (81.2 percent for whites versus 71.6 percent for nonwhites). Differences by sex are not nearly as large, however. Males are only slightly more apt to report an SSN (81.0 percent for males as opposed to 79.5 percent for females). Figure 4 clearly illustrates this marked disparity in reporting rates between the races and the only marginal difference between the sexes.

Figure 4.—SSN REPORTING BY RACE AND SEX



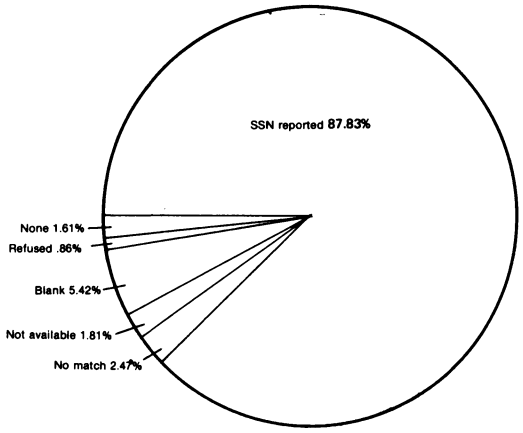
Both race and sex exert some influence on whether persons respond with the answer "none." As expected, females were slightly more likely than males to answer "none" (4.9 percent as opposed to 2.8 percent) and nonwhites were more likely than whites (5.9 percent compared to 2.6 percent). Race, but not sex, seems to determine, to some extent, whether the SSN question is left blank or filled in with "not available." Blank entries were approximately 13.0 percent for nonwhites and, for whites, only nine percent. The percentage of "not available" answers for nonwhites (5.5 percent) is about twice the number for whites (2.7 percent).

Differences in the age distributions of whites and nonwhites account for some of the reporting patterns that have been remarked on. It is fair to say, though, that, for much of the disparity, we must search for other causes.

Individual's Respondent Status.--Because one potential source of distortion in the reporting of SSN's does not exist for those individuals who responded directly to the interviewer's question, it is of interest to compare the response rates for all persons with those for the household respondents alone. 6/ We would, of course, expect respondents to be more complete and more accurate about themselves than about others, and, in fact, this seems to be the case.

The SSN reporting rate among respondents for each age-race-sex category is much higher than the reporting rate for all persons in that same category. For all age-race-sex groups combined, about 87.8 percent of the household respondents reported an SSN (see figure 5), as opposed to 80.2 percent for all persons. 7/

Figure 5.—RESPONSE TO SSN QUESTION FOR HOUSEHOLD RESPONDENTS ONLY



When comparing the response rates between the set of all persons and the subset of household respondents, it should be noted that the age distribution between the two groups differs radically. This is because of a reluctance by Census interviewers to accept younger members of households as household respondents. Slightly more than 18 percent of all persons 14 years or more in the CPS are 14 to 20 years old; however, less than four percent of the household respondents were from this age group. Even more extreme are the figures for persons 14 to 17 years old. They represent 11.1 percent of all the adults in the survey but are respondents only 0.8 percent of the time. Given this disparity in the age distribution, a sensible approach to comparing respondents and nonrespondents is to look at them by age, as in the following table:

Table 1.--Proportion Reporting An SSN

	Respondent	Nonrespondent
Total	87.8%	74.4%
14 to 17 year-olds	61.0%	45.5%
18 years or older	88.0%	81.1%

The improvement in reporting is pronounced for the 14 to 17 year-olds. One of the possible reasons for this may be that the older members of the household are not reliable sources of the SSN for the younger family members.

Replies of "none" occur about 1.6 percent of the time for respondents; the proportion for nonrespondents climbs to 5.6 percent. Virtually all of this difference is due, however, to differences between the age structure of respondents and nonrespondents, as can be seen from the following breakdown:

Table 2.--Proportion Reporting "NONE"

	Respondent	Nonrespondent
Total	1.6%	5.6%
14 to 17 year-olds	23.0%	23.2%
18 years or older	1.5%	1.5%

As we expected, the number of blank entries was lower among respondents than among nonrespondents (5.4 percent versus 12.2 percent). The younger groups had a large influence on this; the incidence of blanks among all persons 14 to 20 years of age is 19.8 percent, whereas the incidence of blanks among respondents of the same age is 9.1 percent.

The percentage of answers of "not available" was also lower for respondents than for nonrespondents (1.8 percent as opposed to 3.9 percent). However, the incidence of refusals (0.9 percent for

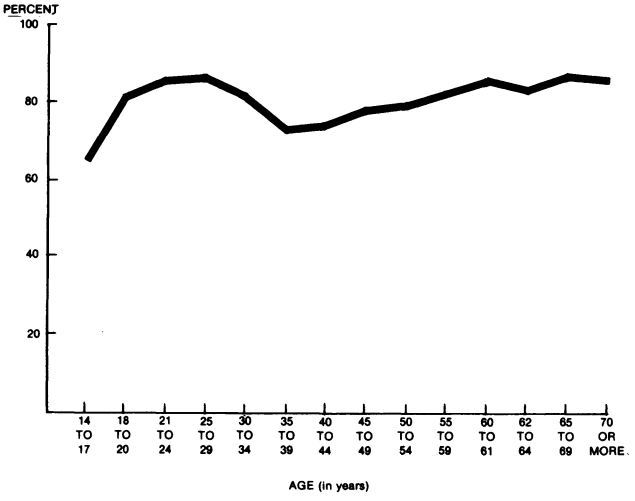
respondents compared to 1.1 percent for nonrespondents), and of nonmatches (2.5 percent versus 2.9 percent), do not represent significant differences.

4. HOUSEHOLD REPORTING BY THE CHARACTERISTICS OF THE HOUSEHOLD RESPONDENT

We have seen in section 3 that the respondent status was a strong determinant affecting the SSN reporting for individuals. In this section, we attempt to analyze how the SSN reporting of the entire household varies according to selected attributes of the respondent.

Household Reporting of SSN's by Respondent's Age.--Figure 6 shows a graph of SSN reporting for all persons based on the age of the respondent. As can be seen, the proportion of SSN's obtained for

Figure 6.—SSN REPORTING FOR ALL PERSONS BY RESPONDENT'S AGE



the household is noticeably smaller when the respondent is either 14 to 17 or 35 to 44 years old. The percentages of SSN's obtained are 65.4 percent and 73.8 percent when the respondent is 14 to 17 years old, and 35 to 44 years old, respectively, versus 82.5 percent for the other age groups. This variation is caused in part by the nature of the households for which these particular age classes were respondents; i.e., these households contained a higher proportion of individuals between the ages of 14 and 17, when they are less likely to have SSN's. In fact, when these age groups (14 to 17 and 35 to 44) are household respondents, the whole household shows a much higher incidence of answers of "none" (8.1 percent and 7.0 percent compared to 3.0 percent for the other age groups).

As mentioned before, the CPS interviewer is instructed to interview the household head or the spouse of the household head, if possible. Therefore, when a younger member of the household is chosen as a respondent, it is generally because older members are not at home at the time of the interview. However, additional household members may be present when the head of household is the respondent. This situation may account for the high incidence of "not available" responses for households with 14 to 17 year-old respondents (6.2 percent versus 2.9 percent for the other ages). Blank entries are also more prevalent among households having a respondent who is 14 to 17 years of age (15.0 percent as opposed to 9.2 percent overall).

Just as an individual over the age of 35 was found to be slightly more likely to refuse to report his number, respondents over the age of 35 are also slightly more likely to refuse to provide numbers for other household members (1.1 percent versus 0.6 percent).

Household Respondent's Race and Sex.--As was the case when SSN reporting was examined by the characteristics of the individual, the race of the respondent had a greater influence than did sex in determining the proportion of SSN's reported for the household. About 81.4 percent of the persons 14 years or older in households of white respondents gave an SSN versus 71.9 percent of the households of nonwhite respondents. Whether the respondent was male or female made a difference of only 1.4 percent (reporting rates by sex of the respondent are 81.4 percent for males and 80.0 percent for females).

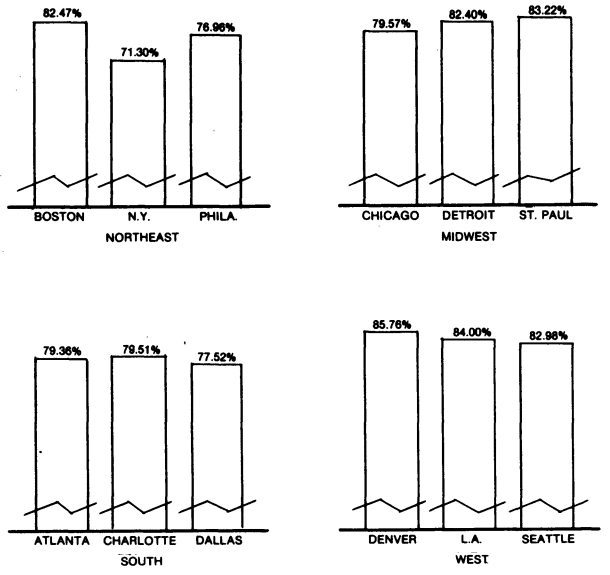
5. OTHER ASPECTS OF SSN REPORTING

Data Collection Centers.--The twelve regional offices of the Census Bureau both train and supervise the CPS interviewers and, thus, exercise some influence on the quality of the data collected. Also, the final manual processing of the control card--that of its being keyed onto tape for computer use--was divided into work units within data collection centers.

The SSN reporting rate does vary between the centers; New York's rate of 71.3 percent is significantly lower than the others, while Denver's performance of 85.8 percent is over five percent better than the overall rate of 80.2 percent for all the centers (see figure 7).

The low reporting rate of New York is due, in part, to a much higher proportion of blanks (13.7 percent versus 8.9 percent when New York is excluded) and almost twice as many "not available" responses as the other centers (5.5 percent as opposed

Figure 7.—SSN REPORTING BY REGIONAL DATA COLLECTION CENTER



to 2.8 percent). The somewhat lower reporting rates of Philadelphia and Dallas, at 77.0 and 77.5 percents, respectively, also reflect slightly higher blank and "not available" entries. In addition, all three centers have higher proportions of answers of "none" to the SSN question than do the other centers.

Rotation groups.--Examining the data by rotation groups, we find there is a direct relationship between the incidence of SSN reporting and the number of months that the rotation group was in the survey during the time frame that the SSN question was asked. (This was from December 1972 to March 1973.) The table below illustrates this association.

Table 3.--SSN Reporting Rate by Number of Months SSN Question Asked

Rotation Groups	Months Question Asked	SSN Reporting Rate
1 and 5	December-March	82.8
2 and 6	January-March	81.8
3 and 7	February-March	78.7
4 and 8	March only	77.5

Rotation groups 1 and 5 were eligible for interview during all four months and show the highest response rates. The two rotation groups (4 and 8) which were in the survey for only one of the four months have the lowest SSN reporting rates.

Conclusion.--The steps taken by the Census Bureau in conducting the survey are seen to be carefully controlled and administered; nevertheless, chances for human and machine error exist. Although not an exhaustive study of all the possibilities, we have in this paper identified some variables affecting the likelihood of SSN's being reported in the CPS. Age is a very strong factor, and race, more so than sex, plays an important role in determining whether a number is provided. If an individual was the household respondent, chances are improved that his number will be reported. Also, the characteristics of the respondent affect to some extent the reporting rate of his entire household.

FOOTNOTES

1/ Detailed information on the Current Population Survey is available in many Census Bureau and Bureau of Labor Statistics publications. A particularly good source on the CPS as redesigned after the 1970 Census is a paper in the Annals of Economic and Social Measurement, "The Current Population Survey: an overview" by Marvin Thompson and Gary Shapiro (vol. 2, 1973, pp. 105-119).

2/ Linkages between the survey and administrative records have been attempted during most of the years since 1963. However, unlike the 1973 Match Project, these earlier linkage attempts were made only with a subsample of the cases in the survey. In some instances, like the Pilot Linkage work referenced in the text, the subsample was representative of the total civilian noninstitutional population. In other instances, only certain subgroups were represented, such as persons 60 years or older.

3/ It was possible, as is described in the second paper at the session, to find about half of the SSN's which were not reported initially. SSN's found for persons who refused, however, could not be returned to the Census Bureau, since to do so would violate the respondent's expressed wish in this regard. Instead, at Social Security, the characteristics of the refusals are being analyzed as a group and adjustments to the overall match results will be made accordingly.

4/ Our best estimate is that about 98 percent, or perhaps a little more, of the U. S. noninstitutional population 17 years or older have been issued SSN's. For persons 14 to 17 years old, between 80 and 90 percent have been issued numbers.

5/ Age is determined as of March 17, 1973. This differs from the age variable used

in the other two papers, which is in terms of age attained in 1972.

6/ The March household respondent is being identified here. Usually, but not always, this was the same individual who responded to the SSN question. The SSN information may have been obtained at any time from December 1972 through March 1973, depending on the rotation group in which the household was a part.

7/ Recall that, in the discussion of previous studies, it appeared that our SSN reporting rate improved to about 80 percent over the 1966-67 Census study rate of 76 percent. It would be more valid to contrast our 87 percent SSN reporting rate for respondents, since each person in the 1966-67 study was to respond for himself.