RECORD CHECK STUDY OF ACCURACY OF INCOME REPORTING

Introduction

This report presents an evaluation of the accuracy of income data compiled in the 1960 Census of Population. This evaluation is based upon a comparison of income data obtained from a sample of Internal Revenue Service (IRS) tax returns with the income data reported by the identical persons in the 1960 Census. In addition to an evaluation of accuracy, a secondary objective of this report is a comparison of differences in income distribution covering not only persons who have reported to both sources but also persons who have reported to only one of the sources. This information is used to develop relationships between IRS and census data by different income types, by type of reporting units, and by persons reported in both or only one of the data sources.

Measurement errors on census items can arise from a number of different sources; e.g., people missed by interviewers will result in undercount, personal characteristics may be erroneously reported, people may fail to report some of the information requested of them, adjustments for the missing data may introduce other errors, etc. This report investigates only one of these sources of possible errors, i.e., "erroneous" reporting. It analyzes differences between census income data as compared with IRS adjusted gross income data (less capital gains and losses) with respect to specific types of income and income reporting units. There are, however, basic differences in coverage and definitions between census and IRS data; consequently, the differentials will have to be interpreted selectively.[1]

The procedure used in this report is described as a matching study because it compares, for the same individuals, income as reported to the Internal Revenue Service for a sample of persons with income in the census records. All work is done by the Bureau of the Census in order to preserve the confidentiality of replies to census questions; no one other than Bureau employees, who are sworn to uphold the confidentiality of all census information, has access to the census returns used in the comparisons. The confidentiality of IRS records was also safeguarded. The only products of such studies are statistical tables summarizing the characteristics sought.

In this particular matching study, income information for tax year 1959 was initially obtained from a subsample of individual tax returns selected on a probability basis from the IRS Statistics of Income sample. Census household schedules for the IRS sample persons were located, and information on the composition of each household was transcribed. Additional tax returns were requested for those household members 14 years old and over who should have filed an IRS tax return. For those for whom tax records were found, pertinent information from the census schedules was transcribed and was subsequently matched with previously transcribed IRS information.

[1]See further discussion on this point on page 7.

Data Presented

Three types of comparative data are presented. The first type shows absolute levels. Tables 1 to 8 cover distribution of census income data cross-classified by IRS adjusted gross income data. Tables 9 to 12 show amounts of discrepancy between census and IRS income data. Tables 13 and 14 show the income distribution (IRS and census income class) by farm and nonfarm residence of persons filing IRS Schedule F (Schedule of Farm Income and Expenses).

The second type of comparative data shows percentage distributions. To facilitate analysis, percentage income distributions of census and IRS data by specific types of income and by census reporting units are shown in summary tables A and B. These income types cover total money income, wage or salary income, self-employment income, and income other than earnings.

The third type of comparative data includes a number of standard indexes that summarize the magnitude and directions of response differences and is shown in tables 15 to 17.

Statistics relating to small subgroups of the population or rare sources of income are subject to relatively large sampling errors and the reader should keep this in mind when analyzing the results. Table D (p. 7) indicates the size of the sampling error of estimates in this report.

Summary of Findings

Overall findings indicate that census income distributions, especially for total money income and for wages and salaries, are generally comparable to IRS income distributions. Gross differences within the income class interval matrix offset each other. Findings are shown in tables A and B and 1 to 14. As shown in summary table A, the mean total money income of married persons filing jointly as derived from census data is about 3 percentage points higher than that obtained from the IRS data. For mean wage or salary income, the difference is also about 3 percentage points, but in the opposite direction. For self-employment income and income other than earnings, the mean incomes derived from the census are higher than the IRS means by 43 percent and 48 percent, respectively. These relatively large differences for the latter two groups, however, have only a slight impact on the net difference for mean total income because wages and salaries comprise the bulk of aggregate income reported--about 76 percent according to the census and 81 percent according to IRS records.

The small difference in wage or salary income is understandable since this particular income type can be estimated by respondents without much difficulty. Wages and salaries are received regularly, and records such as tax withholding statements are available for reference purposes.

The relatively large difference in self-employment income (comprised of both nonfarm and farm self-employment income) can be partly explained by the following reasons:

First, net farm self-employment income reported on tax returns tends to be smaller than that reported in other data sources because of differences in coverage and definition of items reported among these sources.[2] As shown in the second column of table 13, mean self-employment income of persons with income reported on IRS Schedule F (Schedule of Farm Income and Expenses) by persons reporting farm residence in the census was about $900. Data shown in the second column of table 14 show census mean self-employment income of about $2,300 for persons with income reporting farm residence. If it is assumed that all of this amount was derived from farm self-employment income, then farm self-employment income reported in the census was more than double the mean net farm income reported in the IRS Schedule F (Schedule of Farm Income and Expenses).

[2]See F. D. Stocker and J. C. Ellickson, "How Fully Do Farmers Report Their Incomes?," *National Tax Journal*, June 1959, pp. 116-126; H. M. Groves, "Empirical Studies of Income Tax Compliance," *National Tax Journal*, December 1958, pp. 297-301; and H. S. Houthakker, "The Great Farm Tax Mystery," *Challenge*, Jan.-Feb. 1967, p. 12.

Second, there may have been response errors in the census such as the possible reporting of gross income instead of net income, understatement of losses or self employment operational expenses, or possible inclusion of rental income in self-employment income instead of the "income other than earnings" category. Table A shows that 4.9 percent of joint returns indicated "losses" for self-employment income as contrasted with only 0.3 percent for reporting units in the census.

The higher census mean for "income other than earnings" reported by married persons filing joint returns can be partly explained by the fact that certain types of transfer payments, e.g., Social Security payments and veterans' pensions, are included in the census but are not reported in IRS tax returns. Data in table A show that 38 percent of the census units reported some amount of "income other than earnings" as compared with 30 percent of IRS units. Also, relatively more census units (35 percent) than IRS units (27 percent) were covered in the $1 to $2,999 income interval. These data indicate that census units reported "income other than earnings" amounts which were not included in the IRS returns. Table A also indicates that the income distribution obtained from IRS joint returns and census married-person units has a higher degree of agreement than the

Table A.--**Percentage Distribution of Married Persons in the 1960 Census Filing IRS Joint Returns, by Income Reported in the Census and on IRS Joint Returns, by Income Type, 1959**

(Numbers in thousands. Percents may not add to 100.0 due to rounding)

Income classification	Total money income		Wage or salary income		Self-employment income		Income other than earnings	
	Census	IRS	Census	IRS	Census	IRS	Census	IRS
Total	36,525	35,743	36,525	33,944	36,525	33,648	36,525	35,667
Census units with "income not reported"[1]	782	-	2,581	-	2,877	-	858	-
Census units reporting both IRS and Census income	35,743	35,743	33,944	33,944	33,648	33,648	35,667	35,667
Total reported	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
None	2.8	-	11.2	11.1	78.5	75.3	61.7	70.1
Loss	0.1	1.0	-	-	0.3	4.9	-	0.1
$1 to $599	1.1	1.5	2.3	2.8	3.0	3.8	20.7	20.2
$600 to $999	1.0	1.9	1.9	1.3	1.6	2.1	6.4	2.7
$1,000 to $1,999	5.3	7.5	4.0	5.6	3.4	3.7	6.4	3.3
$2,000 to $2,999	7.6	9.4	8.0	7.4	2.1	2.8	1.5	1.1
$3,000 to $3,999	9.5	9.2	7.7	6.7	2.4	2.2	1.3	0.8
$4,000 to $4,999	13.8	13.3	13.1	13.0	1.8	1.1	0.5	0.6
$5,000 to $5,999	12.7	12.1	12.5	12.0	1.6	0.6	0.2	0.5
$6,000 to $6,999	13.0	11.8	12.4	12.0	0.9	0.7	0.4	-
$7,000 to $7,999	8.3	9.1	7.2	8.4	0.9	0.1	0.1	-
$8,000 to $8,999	6.1	6.0	5.4	5.8	0.7	0.4	0.1	0.1
$9,000 to $9,999	4.6	4.3	4.5	4.4	0.2	0.4	0.1	0.1
$10,000 to $14,999	9.7	9.0	7.5	7.8	1.1	0.7	0.4	0.1
$15,000 to $19,999	2.3	2.1	1.3	1.1	0.8	0.6	0.1	0.1
$20,000 to $24,999	0.8	0.4	0.4	0.2	0.4	0.2	-	-
$25,000 and over	1.5	1.4	0.5	0.6	0.6	0.5	0.1	0.1
Mean income....dollars..	6,611	6,414	5,321	5,465	1,094	763	525	355
Net percent difference[2]	+3.1	(X)	-2.6	(X)	+43.4	(X)	+47.9	(X)
Aggregate income:								
Census....billion dol..	236.3	(X)	180.6	(X)	36.8	(X)	18.7	(X)
IRS....billion dol..	(X)	229.3	(X)	185.5	(X)	25.7	(X)	12.7

- Represents zero; minus sign (-) before a figure denotes decrease. X Not applicable.

[1]See page 7 of the text for discussion of this item.

[2]$\frac{\text{Census} - \text{IRS}}{\text{IRS}}$ (100).

income distribution obtained from IRS individual returns, and census unrelated individuals as shown in table B.

As summary table B shows, census mean total money income of fully matched families in which all members were accounted for on IRS tax returns was about 2 percentage points higher than the comparable figure obtained from IRS data.[3] For all families, including both fully matched and partially matched families, census mean total money income was about 14 percent higher than the comparable IRS income data. This difference is due primarily to inclusion of household members 14 years or over reporting income in the census but excluded from IRS data. For unrelated individuals, the census mean total money income was higher by 7 percentage points than the mean income obtained from IRS data.

Data in tables 9 to 12, which show the direction and extent of gross underreporting or overreporting of census income data, relative to data reported on tax returns, reveal relatively large gross differences; but overall, these offset each other. For example, in table 9, which shows the amount of discrepancy when census total money income for married persons was greater than IRS total money income for joint returns, approximately 3 million returns, or 18 percent of the 16.6 million returns had a discrepancy of $2,500 or more. On the other hand, of the total 17.6 million returns having census income less than IRS income, approximately 2.4 million returns or 14 percent had a discrepancy of $2,500 or more.

The standard indexes of response differences shown in tables 15 to 17 provide convenient summaries of the effect of response errors. Response errors include both response variance and response bias. The response variance tends to cancel out when the number of observations is very large. The remaining bias reflects effects of systematic differences that are consistent in one direction. A number of different measures are shown which are described in detail in the next section. Findings relating to some of these measures are summarized below.

The index of inconsistency is a measure of response variance. The higher the index of inconsistency, the less reliable is the measurement of a specific characteristic.

[3]For a description of matched, partially matched, and unmatched families, see page 7 of this report.

Table B.--Percentage Distribution of Selected Census Units, by Income Reported in the Census and on IRS Tax Returns, by Income Type, 1959

(Numbers in thousands. Percents may not add to 100.0 due to rounding)

Income classification	Fully matched families--total money income		All families--total money income		Unrelated individuals--total money income		Unrelated individuals--wage or salary income	
	Census	IRS	Census	IRS	Census	IRS	Census	IRS
Total	35,050	34,396	41,236	40,583	6,690	6,200	6,690	5,296
Census units with "income not reported"[1]	654	-	654	-	490	-	1,394	-
Census units reporting both IRS and Census income	34,396	34,396	40,583	40,583	6,200	6,200	5,296	5,296
Total reported	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
None	2.4	-	2.0	-	3.9	-	4.9	7.8
Loss	0.1	0.9	0.1	0.8	-	1.7	-	-
$1 to $599	1.2	1.5	1.0	2.4	5.8	8.4	9.4	7.2
$600 to $999	1.0	1.9	0.9	2.4	10.7	12.7	8.2	10.6
$1,000 to $1,999	4.4	7.1	4.2	8.2	16.4	20.5	19.5	19.7
$2,000 to $2,999	7.0	8.3	6.7	9.0	19.5	16.2	15.4	15.6
$3,000 to $3,999	9.9	9.7	9.8	10.1	10.7	13.3	10.3	11.8
$4,000 to $4,999	12.2	12.9	11.9	12.4	11.8	11.6	15.0	13.6
$5,000 to $5,999	12.7	11.6	11.8	11.1	10.3	7.5	9.1	7.6
$6,000 to $6,999	13.2	10.9	12.9	10.2	4.8	3.5	5.5	3.8
$7,000 to $7,999	8.2	8.4	8.4	8.2	3.2	1.8	1.5	0.2
$8,000 to $8,999	6.1	6.2	6.7	6.0	0.5	0.9	-	0.8
$9,000 to $9,999	5.3	5.2	5.5	4.6	0.4	0.6	-	-
$10,000 to $14,999	11.3	10.9	12.5	10.5	1.4	1.0	0.8	0.8
$15,000 to $19,999	2.7	2.6	2.9	2.3	0.7	0.4	0.5	0.5
$20,000 to $24,999	0.9	0.4	0.8	0.4	-	-	-	-
$25,000 and over	1.4	1.4	1.7	1.3	0.1	0.1	-	-
Mean income.....dollars..	6,812	6,663	7,300	6,408	3,202	2,991	2,898	2,770
Net percent difference[2]	+2.2	(X)	+13.9	(X)	+7.1	(X)	+4.6	(X)
Aggregate income:								
Census.....billion dol..	234.3	(X)	296.3	(X)	19.9	(X)	15.3	(X)
IRS.....billion dol..	(X)	229.2	(X)	260.1	(X)	18.5	(X)	14.7

- Represents zero. X Not applicable.

[1]See page 7 of the text for discussion of this item.

[2]$\frac{\text{Census - IRS}}{\text{IRS}}$ (100)

NOTE.--Mean and aggregate incomes based on IRS returns and census units reporting income.

Also, this measure is related inversely with the percent of IRS class and census class identically reported. Thus, as the index of inconsistency rises, the percent identically reported in both sources declines. The data show that the index of inconsistency for wages and salaries was the most consistent among the indexes for the different sources of income. Self-employment income was more consistent than "income other than earnings."

The net difference is an estimate of bias. If one of the sources of information used in the matching procedures is mostly free from error, it can be considered a standard. The IRS data cannot be uniformly considered a "true" standard because the coverage of income recipients and the income definitions are not identical with those of the census. However, as noted before, there are different degrees of comparability among different income types. Thus, wage or salary income data from the IRS and the census tend to correspond fairly well.

The net difference rate as used here does not provide an estimate of the amount of "bias," but of the degree of variation found between information obtained from the two sources. When this rate is negative, the number of census units reporting is less than the IRS units reporting in a specific income interval. Data in table 15 show that for total money income, the net difference rates were generally negative in intervals under $3,000 but positive (more census units reporting relative to IRS units) in intervals over $3,000.

The index of net shift relative to IRS class is the ratio of the net difference between IRS and census units to the number in a class reported in the IRS. Among the three income sources, the index of net difference is least for wage or salary income.

Indexes of Response Variance and Bias

The response errors of a particular census or sample survey arise from the combined effects of response bias and response variance. Measures of these two items can therefore be used as indexes of the accuracy of the data. Response bias represents systematic errors in reporting data or the effect of types of errors consistent in direction and that would be consistent in independent repetitions of the survey under the same general conditions. Response variance, on the other hand, is the effect of errors that cancel out when a large number of observations are made. A more complete description of these terms follows.

Under certain fairly general survey conditions, matched information provides estimates of response variance; and, to the extent that one of these sources is based on more adequate measurement methods and is acceptable as a standard, it can also provide estimates of bias. Various measures of response variance and bias can then be constructed from the results of this kind of match. Comparison of IRS data with the census gives two measurements for sample persons and for families for each type of income by item and roughly satisfies the conditions given above. A group of such measures, which appears to be useful for analytic purposes, has been computed for each income item.

Table C illustrates the results of the comparison of census data with IRS data where the value 1 is assigned to a person classified as reporting an amount (including zero) of a specified income type and the value 0 otherwise. (Persons who reported no income information for a specified income type are excluded.) Table C shows that "a" of the persons were classified as having the specified income in both the census and IRS data; "a + c" were classified as reporting the income in the census, and "a + b" were classified as reporting the income in the IRS data.

Table C.--**Representation of Results of Census and IRS Information for Identical Persons**

Results of the IRS	Results of Census		
	1	0	Total
1.....	a	b	a + b
0.....	c	d	c + d
Total...	a + c	b + d	n=a+b+c+d

If x_i represents the result for a person in the census and y_i represents the result for that same person in IRS data, the response difference, which is either 0, +1, or -1 for that particular person, is represented as $x_i - y_i = e_i$. The sum of the values of e_i over all persons included in both the census and the IRS is the net difference. In the notation of the diagram

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = (a+c) - (a+b) = c - b.$$

The gross difference can be represented by b + c. The values of a, b, c, d, the gross difference, and the net difference are the components of the indexes of response variance and bias.

In evaluating a census statistic, the mean square error (MSE) of that statistic is of particular interest. The components of the MSE are as follows:

$$MSE_{x_c} = \sigma^2_{x_c} + B^2_{x_c}$$

where $\sigma^2_{x_c}$ is the variance of the census statistic and $B^2_{x_c}$ is the square of the bias of the census statistic. (Generally, the MSE is defined as having the sampling variance as a third component. For a complete census, the sampling variance vanishes. Even though the items analyzed here were sample items in the census, the sample at the national level was so large that the sampling variance is a trivial part of the MSE. For statistics for small cells or small areas, the sampling variance contribution may be important. The analysis in this report relates to national statistics.)

For data on income, especially wage or salary income, the expected value of the census result minus the expected value of the IRS result is equal to the bias of the census statistic, since responses to the Internal Revenue Service are considered to be more accurate than those reported

in the census. The estimated variance of the individual response differences is

$$s_e^2 = \sum_{i=1}^{n} \frac{(e_i - \bar{e})^2}{n-1}$$

where e_i is the response difference and $\bar{e} = \frac{\sum e_i}{n}$ represents the mean of the response differences. Since $e_i = x_i - y_i$,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (x_i - y_i)^2.$$

Whenever the responses in the census and IRS data are different, $e_i^2 = 1$, since $(x_i - y_i)^2 = (1)^2$ or $(-1)^2$. Whenever the responses are the same, $e_i^2 = 0$. Therefore, $\sum_{i=1}^{n} e_i^2 = b + c$, the sum of all the differences in response from the census and IRS data, or the gross difference. Now, since $\sum_{i=1}^{n} e_i = c - b$, s_e^2 can be written as follows:

$$s_e^2 = \frac{b+c}{n-1} - \frac{(c-b)^2}{n(n-1)}.$$

The gross difference can be expressed as

$$b + c = (n-1)\, s_e^2 + \frac{(c-b)^2}{n}.$$

The gross difference rate is then

$$\frac{b+c}{n} = \frac{(n-1)}{n} s_e^2 + \frac{(c-b)^2}{n^2}$$

The indexes which are described more fully below are functions of the detail in classification of the characteristic. For example, the tables on income presented in this report are in terms of a set of income classes. If the data were tabulated by smaller (or larger) income classes, the indexes would change.

The indexes computed by the formulas shown below generally result in numbers between -1 and +1. In this report, the indexes have been expressed as percents, i.e., the results of the computations have been multiplied by 100.

1. Gross difference rate:

$$g = \frac{b+c}{n} = \frac{(n-1)s_e^2}{n} + \frac{(c-b)^2}{n^2}$$

When n is large, the first component of the gross difference rate is approximately equal to the simple response variance of the census statistic when the difference between the IRS data and the census is used as a measure of the bias. The second component is the square of the estimated bias of the census statistic. If the bias is small, the gross difference rate can be used as a measure of the simple response variance of the response difference.

2. Index of inconsistency:

$$\hat{I} = \frac{g}{2pq} = \frac{g}{p_1 q_1 + p_2 q_2}$$

This index shows the ratio of the simple response variance g/2, to pq where p is the average proportion in the census and IRS data having the specified characteristic. An estimate of pq is $\frac{p_1 q_1 + p_2 q_2}{2}$; $p_1 = \frac{(a+c)}{n}$ is the proportion of matched persons in the IRS sample having a specified characteristic in the census;

$$p_2 = \frac{(a+b)}{n}$$

is the proportion of matched persons in the IRS sample having the same characteristic in the IRS,

$$q_1 = 1 - p_1 = \frac{(b+d)}{n} \text{ and } q_2 = 1 - p_2 = \frac{(c+d)}{n}.$$

Therefore, $\hat{I}$ is estimated in the following way:

$$\hat{I} = \frac{(b+c)/n}{\left(\frac{a+c}{n}\right)\left(\frac{b+d}{n}\right) + \left(\frac{a+b}{n}\right)\left(\frac{c+d}{n}\right)}$$

A simple interpretation of $\hat{I}$ is as follows:

Assume that a sample of n elements is drawn with equal probability and with replacement. Also, assume that the between-element covariance of response deviations is zero--that is, that the quality of response of one person is independent of the quality of response for any other person. Then, for a sample of one element, the total variance can be expressed as the binomial variance, pq. The total variance is, then, the sum of the simple response variance and the "pure" sampling variance. Therefore, the simple response variance is equal to or less than pq. As stated above, g/2 is an estimate of the simple response variance.

As the measurement of the specified characteristic becomes less reliable but remains unbiased, the simple response variance increases and the sampling variance decreases. When the measurement process becomes equivalent to tossing the same coin for each element ($0<p<1$ and constant for all trials), the response variance is equal to the total variance. The index of inconsistency is useful in determining the consistency or reliability of a zero-one variate included in the census.

The index of inconsistency as shown lies between 0 and 100, if the assumptions given above hold. However, the estimator of the index can be greater than 100. Such items have been identified in tables 15 to 17. In all cases, the closer the $\hat{I}$ is to 100, the less reliable is the item.

In most evaluative studies, it is difficult to adhere to these assumptions. In particular, the IRS data cannot be considered a repetition of the census procedure. However, the index $\hat{I}$ appears to be a useful indicator in assessing the relative consistency of recorded responses, as between characteristics and as between censuses, even with such deviations from the theoretical assumptions.

3. Net difference rate:

$$\bar{e} = \sum_{i=1}^{n} \frac{e_i}{n} = \frac{c-b}{n}$$

This index gives an estimate of the amount of bias in the census statistic. If the sign is negative, there is an understatement in the census.

4. Index of net shift relative to IRS results:

$$\frac{\bar{e}}{p_2} = \frac{c - b}{a + b}$$

This index shows the ratio of the net difference to the number in the class reported in the IRS data.

5. Percent of population units identically distributed relative to IRS results:

$$r = \frac{a}{a + b}$$

Since IRS information is taken as the standard, this index gives an indication of the stability of the response relative to the standard. This index has an interesting relationship to the index of inconsistency. When the proportion of persons with the specific characteristic in IRS data is small, the two indexes are complementary. When the proportion of persons with the specific characteristic in the IRS data is large, the index of inconsistency provides a more reliable measure of the stability of response. However, "r" appears to be a useful index because its form is simpler than the index of inconsistency. Furthermore, its meaning and implication can be grasped more easily.

IRS-Census Match Survey Methods and Design

Sample Design and Selection

The data presented in this report are based on a sample of about 3,100 tax returns. Initially, 2,300 returns were subsampled from a sample of 10,370 individual income tax returns. Each year the Internal Revenue Service selects a systematic sample of all income tax returns filed for the year. Data included in their report: *Statistics of Income, Individual Tax Returns,* are based on this annual sample of returns, which is stratified by income size. The 10,370 cases represent a subsample of the *Statistics of Income* sample.

The Internal Revenue Service supplied to the Census Bureau a listing by income stratum of all serial numbers of tax returns in their sample. A sample of the serial numbers was selected from each stratum, the effective sampling rate being about 1 in 23,000 for income of less than $50,000 and 1 in 700 for incomes of $50,000 and over.

A photocopy of each return for which the serial number had been selected was requested. A copy was received for all except approximately 900 cases for which the serial numbers had been changed. A scheme was devised to include the missing return cases in the sample at approximately the same rate as noted above. For high income returns, the sample units were selected throughout the entire country. For low income returns, the sample of 1040A forms was drawn from three of the Internal Revenue Service district offices.

Since income was collected for only 25 percent of all households in the 1960 Census, IRS data could be matched with census income in only one-fourth of the cases. Thus, although there were 10,370 IRS returns initially selected for the sample, the number of original cases on which the data in this report are based is approximately 2,300, which was later augmented by another 800 other-household-member tax returns.

Search-Match Procedures

As the photocopies of the 10,370 sample returns were received, the census files were searched in order to locate the census schedules for the sample persons. For about 11 percent of these persons the schedules could not be located. The composition of each census household in which the sample person was located was then compared with those persons listed on the sample IRS return. Where persons 14 years old and over included in census households were not listed on the tax returns, a request for the tax return of these additional persons was made to the Internal Revenue Service. About 3,500 returns were obtained from this request for additional persons. Because the income item was a 25-percent sample item, tax returns of "missing" other household members which were applicable amounted to 800 returns. Thus, the total number of tax returns on which this report is based is approximately 3,100 returns comprised of about 2,300 original sample tax returns and approximately 800 other-household-member tax returns obtained subsequently in the secondary search.

Processing of Material

As the IRS returns were received, a verification operation was set up to identify and adjust for the following types of problems:

1. Many returns were unaudited and some of these had numerical errors.

2. Information was missing from a few of the returns (although most of the returns had complete information).

3. The returns were sometimes illegible.

4. For taxpayers who did not use the printed IRS forms, it was sometimes difficult to determine the actual components of an item.

After verification procedures were completed, the IRS information was transferred onto punch cards, edited, and processed onto a final computer tape.

In a separate operation, pertinent information was obtained from the census records for persons covered in the IRS study. This information was copied onto census transcription sheets, and the transcribed information was verified, edited, and then processed onto a final computer tape.

All documents were handled by Census employees under strict security procedures. Once statistical information was obtained, IRS documents were destroyed. Matching of IRS and census data was made in the computer, producing statistical information only.

Tabulation Process

The tape containing the information from census records was collated (in the computer) with the tape containing the IRS data. Tables were produced covering

(1) married persons filing jointly, (2) persons filing individually, (3) families, and (4) unrelated individuals. Information was obtained on three types of families:

1. Fully matched families--An IRS record was found for all members 14 years and over who reported total money income of $600 or more in the 1960 Census.

2. Partially matched families--An IRS record was found for one or more but not all members (age 14 years and over) of the family who reported total money income of $600 or more in the 1960 Census.

3. Unmatched families--A few households consist of more than one family, in which none of the members of one family is related to any of the members of the other family. Consequently there is a possibility that IRS records were found for members of one family whereas no IRS records were found for the other. Thus, unmatched families represent those families where no IRS records were found for any member of the (secondary) family who reported $600 or more on census schedules.

Treatment of Missing Items

Missing income data were not allocated; that is, values were not assigned to eliminate nonresponses. (In the 1960 Census, when information on income was missing, entries were assigned on the basis of related socio-economic information reported for a similar person.) Instead, in this study, the following procedures were used in tabulating income information. If a person had all census income items blank, it was considered an "income not reported" unit for total money income and for each type. If a person had any income reported, this information was used to obtain the aggregate income values. For example, if census self-employment income was reported as "none" and census "income other than earnings" was not reported, but some wage or salary income was reported for respondent A, his wage or salary income was recorded as his total money income. The same procedure applied to a family unit. Thus, in the above example, if respondent A was the only member of the family who reported on income, the total money income of this family was the wage or salary income reported by respondent A.

However, in the tables showing particular types of income data, e.g., "income other than earnings," the family of respondent A would be classified as a census unit with "income other than earnings not reported." This procedure enables analysis of reporting differences by income types, separately, in addition to "total money income."

Although IRS data are subject to errors of coverage and income reporting, almost all returns included in this study were completely filled out. Where feasible, conceptual differences between reporting of IRS tax returns and census data were adjusted to make the data more comparable; e.g., net capital gains or losses were excluded from IRS data for this study.

Limitations of the Final Data

1. Since the figures in this report are based on sample data, they are subject to sampling variability. The standard errors in table D are measures of sampling variability and apply to the various estimates appearing in this report.

Table D.--**Approximate Standard Errors of Sample Estimates**

IRS incomes under $50,000		IRS incomes $50,000 and over	
Size of estimate	Standard error of estimate	Size of estimate	Standard error of estimate
100,000.....	50,000	2,500.....	1,250
500,000.....	112,000	5,000.....	1,900
1,000,000....	150,000	10,000.....	2,600
1,500,000....	190,000	15,000.....	3,200
2,000,000....	220,000	20,000.....	3,700
3,000,000....	265,000	25,000.....	4,100
4,000,000....	300,000	50,000.....	5,900
5,000,000....	325,000	75,000.....	7,200

2. A basic assumption used to obtain a meaningful interpretation of the net difference is that the IRS tax return provides the standard of accuracy. Although this assumption may be applicable to wage or salary income it may not be applicable to self-employment income and "income other than earnings," because of differences not only in the coverage of persons but also in the definitions used to include or exclude income items. Thus, there are some transfer payments, e.g., Social Security payments, which are reported in the census "income other than earnings" category but which are not reported on IRS tax returns, except under certain conditions.

Consequently, for self-employment income and income other than earnings differences found between census and IRS data should be interpreted as arising from using two different methods of compilation.

3. Schedules for approximately 11 percent of the original IRS sample persons could not be located in the census. This group may have different characteristics; and consequently, the final data, especially in the smaller cells, may be biased due to the exclusion of these persons.

4. Returns for about 20 percent of the additional household members were not located by IRS. There are a number of possible reasons for this, such as failure to file a return and difference in name or address on the IRS return as compared with the census schedule.

5. This study does not include cases in which a household had completed a census schedule but had not filed any IRS return.

6. No search was made for the IRS tax returns of persons 14 years old and over listed as census family members and claimed as dependents on the IRS tax return of persons supporting them. However, a search was made for the IRS tax returns of census family members who were not claimed as dependents on the IRS tax returns.

Related Reports

With respect to coverage and definitions of IRS income data, detailed explanations are found in *The Statistics of Income--1959, Individual Income Tax Returns*, published by the Internal Revenue Service, Treasury Department. Detailed explanation of the census income concepts is given in the text of *U.S. Census of Population: 1960,*

Volume I, *Characteristics of the Population*, Part 1. For a description of the census itself, see *1960 Censuses of Population and Housing: Procedural History*. For the major forms and general description of the procedure used in this study and in other census evaluation studies, see report Series ER 60, No. 1, *Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Background, Procedures, and Forms*. As noted, other reports in the ER 60 Series provide data on other aspects of the accuracy of the census.

A study similar to the 1960 evaluation and research program, the Post-Enumeration Survey, was conducted in 1950. Results of that study are available in the Bureau of the Census, Technical Paper No. 4, *The Post-Enumeration Survey: 1950*, as well as in unpublished memorandums, and in articles published by Census Bureau staff members.

A similar study comparing census and IRS income tax data was also conducted in 1951. A summary of the procedures used and the results of the 1951 study can be found in an article entitled "Income Reported in the 1950 Census and on Income Tax Returns" by Herman P. Miller and Leon R. Paley, included in *Appraisal of the 1950 Census Income Data: Studies in Income and Wealth*, Volume 23, National Bureau of Economic Research, New York, 1958.