FURTHER EXPLORATION OF CPS-IRS-SSA WAGE REPORTING DIFFERENCES FOR 1972

by Beth Kilss and Wendy Alvey*
Social Security Administration

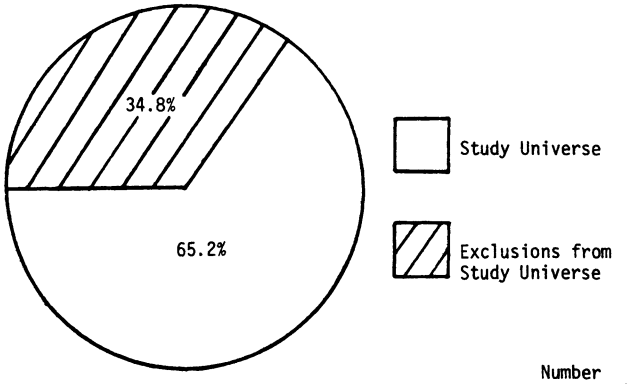
This paper compares wage data from the Census Bureau's Current Population Survey (CPS) with Social Security Administration (SSA) and Internal Revenue Service (IRS) administrative data on wages and salaries for persons eligible for interview in the CPS who filed a nonjoint tax return for 1972. Three-way wage comparisons are made to examine the extent of wage class agreement and evaluate CPS wage reporting. The data base that is used in this analysis was obtained by matching CPS, IRS, and SSA information for the same individual, as described elsewhere in these (and earlier) Proceedings.1/

In the first part of this paper, the study universe and the nature of the exclusions from it are defined. Next, some results obtained by comparing wages from the three sources are presented; and, finally, the study universe is be examined by several variables affecting the quality of wage reporting in the CPS.

RECONCILIATION OF CPS, IRS, AND SSA POPULATIONS

The Current Population Survey is essentially a sample of the civilian noninstitutional population of the 50 States and the District of Columbia. Thus, the population of those who file nonjoint returns includes certain individuals not found among persons eligible for interview in the CPS. These returns must be excluded from the study universe. In addition, because only nonjoint returns with wages from all three sources are being analyzed, further deletions are necessary. Figure 1 briefly summarizes the adjustments made.

Figure 1. --Derivation of Study Universe



	Number (in Millions)
All 1972 Nonjoint Returns (Estimated from 1972 Statistics of Income).....	34.8
Less:	
CPS-IRS Noncomparable Nonjoint Returns.....	2.0
Equals:	
CPS-IRS Comparable Nonjoint Returns.....	32.8
Less:	
Earners with Imputed CPS Wages.....	3.2
Earners with SSA Self-Employment Income.....	0.9
Earners with No Wages from One or More Sources.....	6.0
Equals:	
Study Universe of CPS-IRS-SSA Comparable Nonjoint Returns.....	22.7

As can be seen from the pie chart at the top of figure 1, the study universe is 65.2% of the 34.8 million nonjoint tax returns for 1972. The lower half of this figure provides a list of those that were excluded, the order in which they were excluded, and the size of each. First, approximately two million CPS-IRS noncomparable returns were deleted. These included returns for Armed Forces members, institutionalized individuals, and persons living overseas. Also included were persons filing tax returns for 1972 but who died prior to the March 1973 interview date and persons under 14 years of age. The exclusion of the latter group was necessary, since the March Supplement to the CPS does not ask income questions for them. Also, only 1972 tax-year returns were matched to the CPS; thus, any prior year delinquent returns had to be subtracted out, as well.2/ After these exclusions, 32.8 million CPS-IRS comparable nonjoint returns remained.

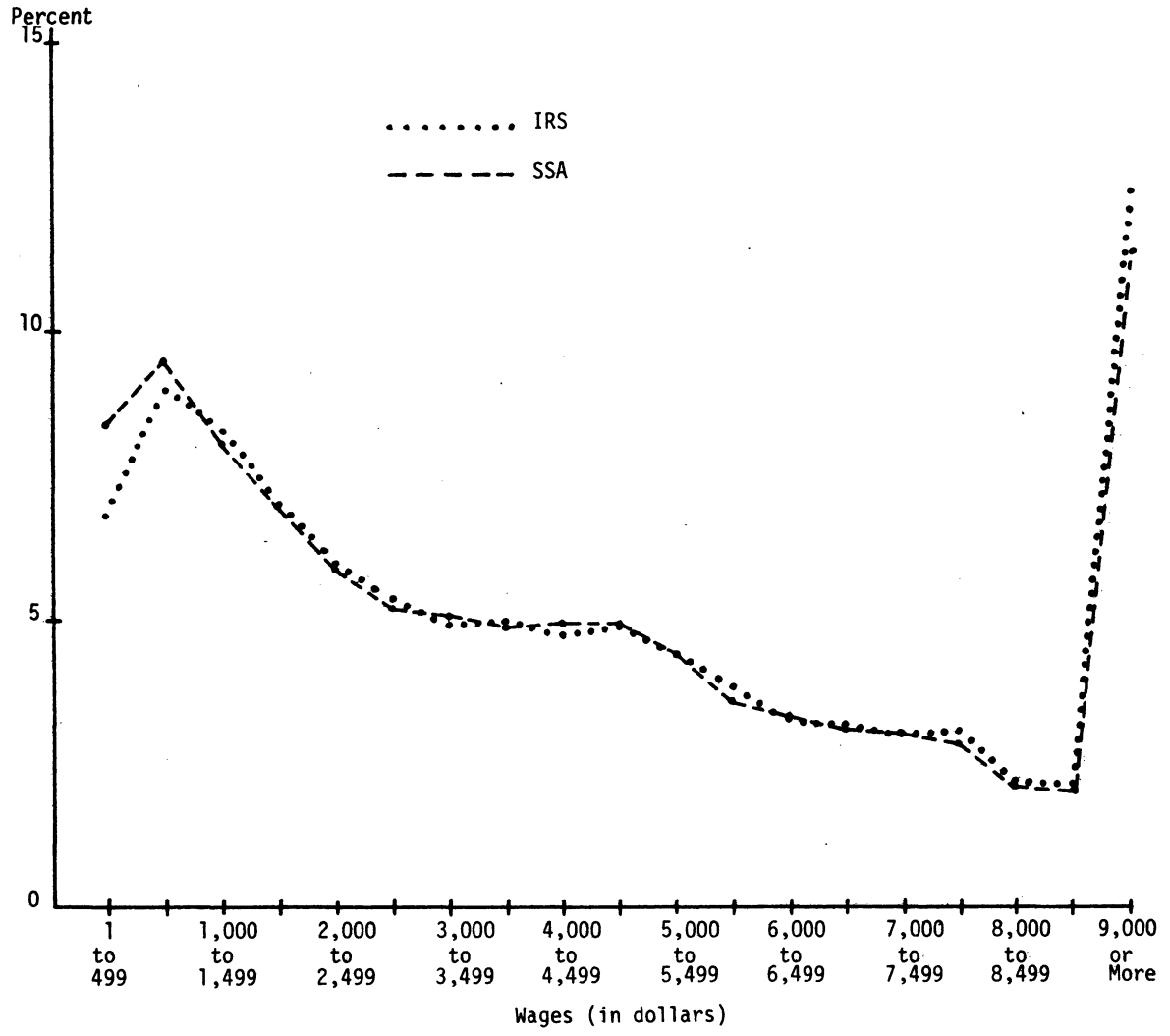
Three other groups were excluded from our analysis.3/ Persons not reporting their wages to the CPS enumerator (and for whom wages were imputed) accounted for the deletion of 3.2 million returns from the study universe. Then, since SSA's computerized administrative files do not distinguish between an individual's wages and self-employment income (this distinction is not required in calculating benefits), about 0.9 million earners with SSA self-employment income were subtracted out.4/ Finally, about six million earners with no reported wages from one or more of the three sources were deleted. This reconciliation yielded a study universe consisting of 22.7 million CPS-IRS-SSA wage earners who filed nonjoint tax returns for calender year 1972. It is this group which is dealt with below.

COMPARISON OF WAGE DISTRIBUTIONS FROM THE STUDY UNIVERSE

Before examining CPS, IRS, and SSA wage class agreement for the same individual, it is interesting to compare wage distributions for the study universe as a whole.

Figure 2 shows the percentage distribution of wages from administrative sources -- IRS and SSA. The 1972 wages are plotted along the horizontal axis in intervals of $500, up to $9,000. Since SSA wages are only reported up to the taxable maximum (which was $9,000 in 1972), for purposes of comparisons among the three sources, it was decided to end the distribution at this point. Therefore, the last class is an open-ended

Figure 2. --IRS and SSA Wage Distributions



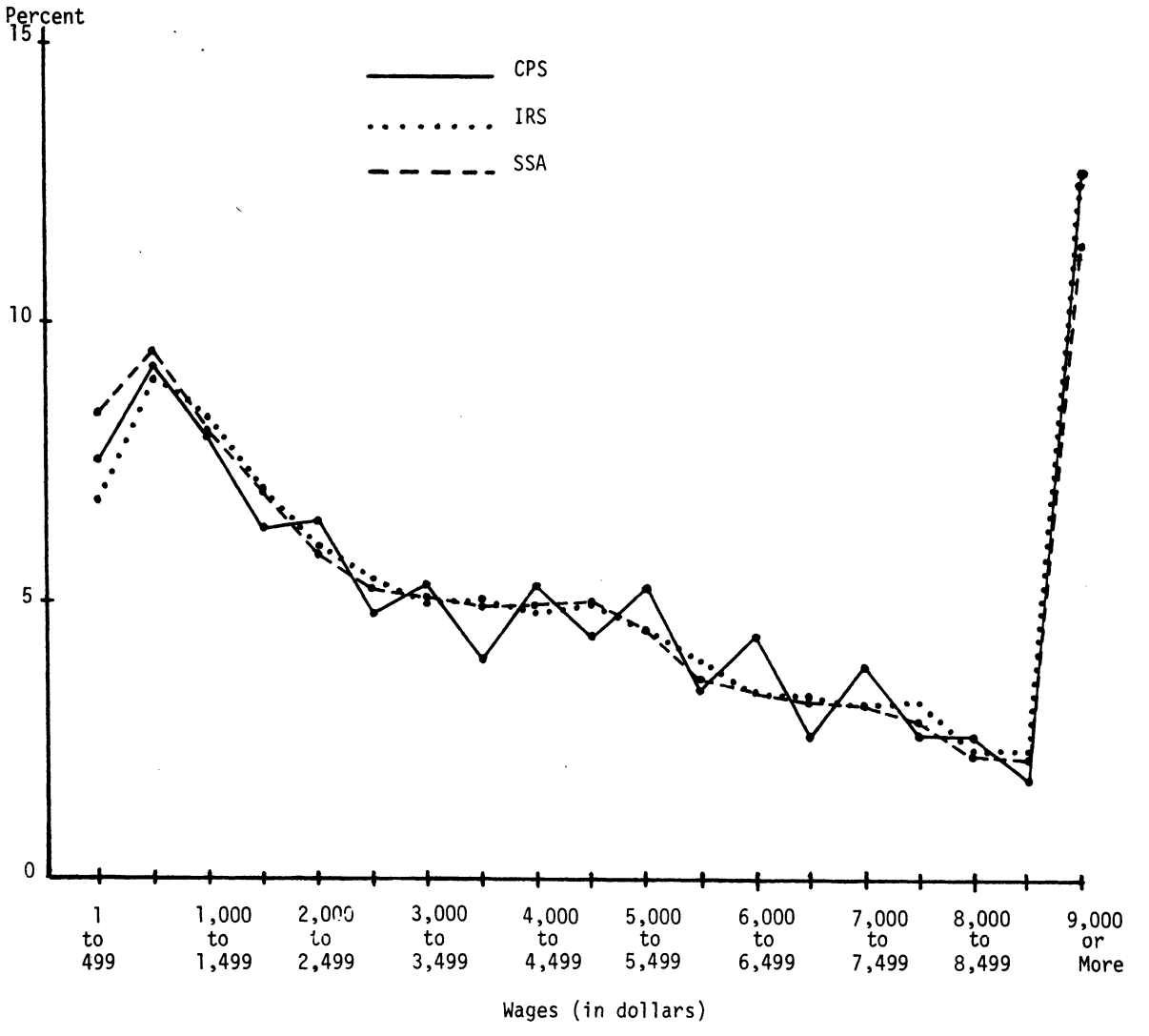
interval of $9,000 or more. The dotted line depicts the IRS distribution; the SSA distribution is shown by the broken line.

For the most part, the reporting of SSA wages appears to be nearly identical to that of wages obtained from IRS data. This was expected since wages from both sources are generally reported from carefully compiled payroll records, with legal penalties imposed on the taxfiler or his employer for fraudulent reporting. For a large part of both distributions -- from $1,000 to $8,999 -- the proportion of wage earners in each class declines as wages increase. However, two main differences in the IRS and SSA distributions are apparent: for persons with less than $1,000, the percent of those with SSA wages notably exceeds the proportion with IRS wages; while, for the $9,000 or more wage class, the reverse is true.

Definitional differences are probably the major factor in these two instances: SSA taxable wages include all wages received by an employee for services rendered in covered employment up to the annual taxable maximum for each employer. However, combinations of covered and noncovered employment are reported on Federal tax returns. It is this, no doubt, which accounts for most of the increased proportion of IRS wage earners beyond the taxable maximum, and, consequently, the smaller proportion at the lower end of the distribution.

Keeping in mind the similarities and differences between these two sources, consider the distribution of CPS wages shown by the addition of the solid line to the distributions presented in figure 2. (See figure 3.) Although the survey data follow the same overall pattern as the administrative data, the distribution is more jagged. This may be attributed to the fact that people often tend to report wages rounded to the nearest thousand dollars [9], thus leading to under- or overstatement of the CPS amount. Also, it is important to remember that wage classes, and not exact wage amounts, are being compared. Therefore, the placement of the upper and lower limits of each interval, as well as the length of each class, all affect what is being observed here.5/

Figure 3. --CPS, IRS, and SSA Wage Distributions



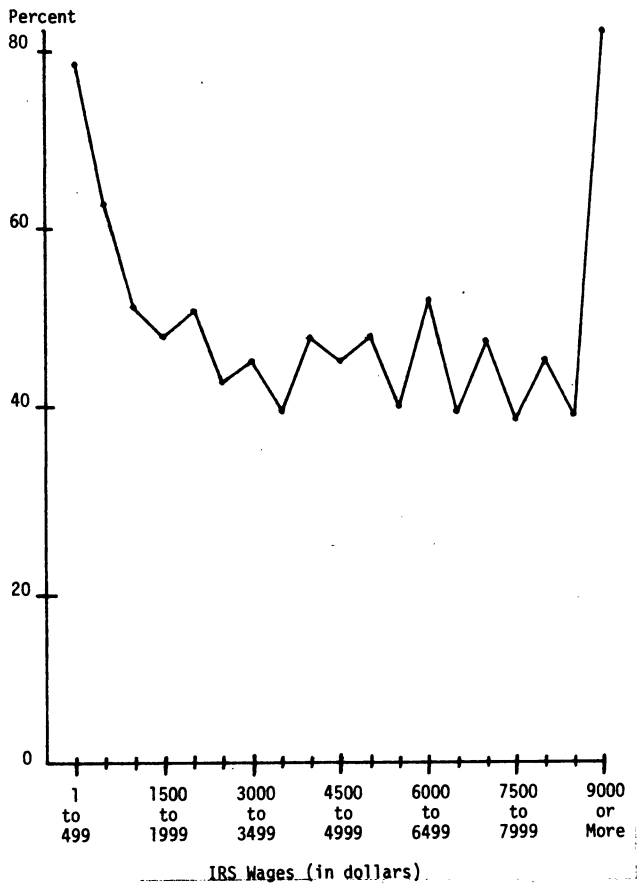
EXTENT OF CPS-IRS-SSA WAGE CLASS AGREEMENT

In the previous section, comparisons were made between marginal distributions of survey and administrative wage data. For purposes of further analysis, CPS, IRS, and SSA wage disparities will be discussed by examining their joint distribution. To do this, wages from all three sources were cross-classified in $500 classes. Agreement was said to exist if wages from more than one source fell into the same $500 interval. Several possible types of agreement resulted: three-way wage class agreement, various combinations of two-way agreement, and cases where none of the wages from any of the sources fell into the same wage class. While tables have been prepared presenting cross-tabulations for all of these agreement types,6/ the analysis in the remainder of this paper will be limited to an examination of three-way CPS-IRS-SSA wage class agreement.

As in figures 2 and 3, the horizontal axis in figure 4 is divided into $500 wage intervals. For the most part, three-way agreement fluctuates between 38 percent and 52 percent for the major part of the distribution. Agreement for the less-than-$1,000 or $9,000-or-greater categories is higher than for the middle-range size classes. This is partly because individuals at the tails of the distribution can only have wage disagreement in one direction. If you will notice, for the wage classes between $1,000 and $9,000, peaks are occurring for the intervals having lower limits which are multiples of $1,000, while the troughs in the graph fall in those classes with lower limits ending in "500." This same sawtoothed effect was observed for the CPS wage distribution in the previous chart.

What are some of the possible factors that may be contributing to these findings? Keep in mind that wages from different sources (administrative and survey) are being compared for the same individual. Because of the very nature of survey data, reporting differences occur. Some of the major "interview factors" which may be affecting CPS reporting are:

Figure 4. --Percent of Three-Way Agreement by IRS Wage Class



1. Type of interview.--manner in which survey information was obtained:

   a. Personal--face to face interview;
   b. Telephone--interviewer phoned individual for answers to survey questions.

2. Response status.--the relationship between the individual answering the survey questions and the individual about whom the survey questions were being asked:

   a. Self-response--the individual answered for himself;
   b. Proxy-response--answers to the survey questions were provided by someone else in the household.

3. Roundedness of CPS wages.--extent to which individual reported exact wage amount in survey interview:

   a. Nonrounded--wage amount ended in a number other than '000' or '500';
   b. Rounded--wage amount ends in '000' or '500'.

These factors are linked to the quality of CPS reporting, and, hence, to the agreement of CPS, IRS, and SSA wages for the same individual.7/ For example, it is expected that an individual who provides wage information from a Form W-2 during a personal interview will be much more likely to have his CPS, IRS, and SSA wages in agreement than a person whose spouse approximated the wage amount in a telephone interview.

In other words, a higher percent of three-way agreement was expected to occur for CPS wage data obtained in personal interviews rather than in telephone interviews, in interviews where the individual responded for himself rather than in cases where the responses were obtained from another household member, and in cases with nonrounded CPS wages rather than with rounded wages.

To test these suspicions, three-way agreement results were examined for each interview factor. The distribution was divided into three sections: less than $1,000, between $1,000 and $8,999, and $9,000 or more.

As shown in figure 5, the expected results occurred for each of the interview factors in the middle-range wage classes; that is, for those with wages of $1,000 - $8,999. However, some differences in the expected agreement pattern occur for wage classes at the extremes.

For example, for wages less than $1,000, three-way agreement is much higher in cases where the individual answered the survey questions for

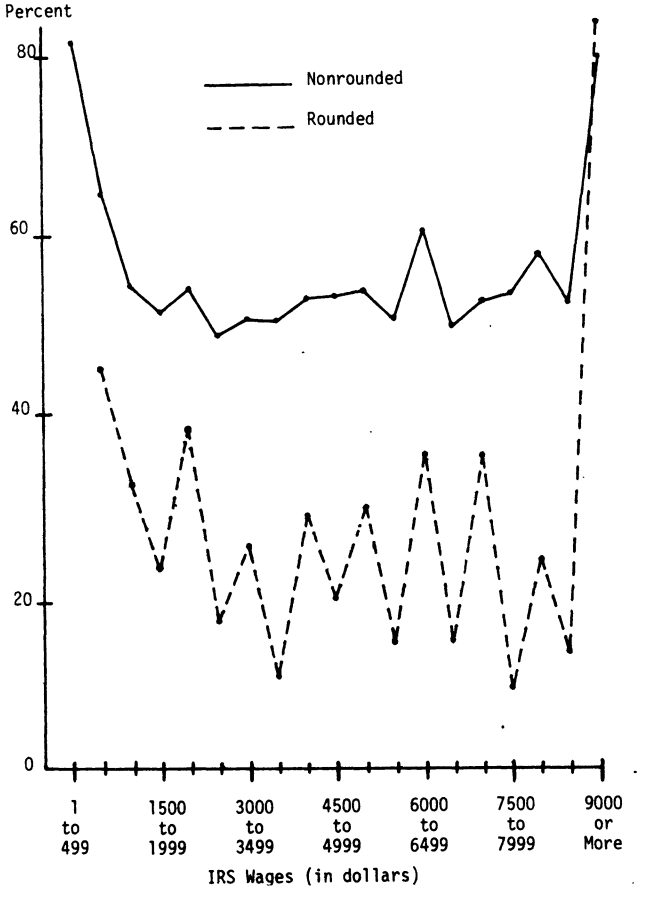
Figure 5.--Percent of Three-Way Agreement by Selected Interview Factors

Selected Interview Factors	IRS Wage Class		
	Less than $1,000	$1,000-$8,999	$9,000 or more
Total.....	61.2	46.4	89.7
PERSONAL INTERVIEW			
Total.....	61.6	47.0	89.8
Self-Response...	50.0	47.8	92.0
Nonrounded.....	55.7	55.2	89.3
Rounded.....	7.5	21.6	95.0
Proxy-Response..	64.2	46.4	86.1
Nonrounded.....	66.4	53.4	88.1
Rounded.....	29.5	24.0	84.2
TELEPHONE INTERVIEW			
Total.....	60.4	45.8	89.5
Self-Response...	46.4	46.8	92.0
Nonrounded.....	48.6	57.0	90.8
Rounded.....	33.0	21.3	93.2
Proxy-Response..	63.9	45.0	83.5
Nonrounded.....	67.6	51.6	82.7
Rounded.....	30.8	27.9	84.1

someone else in the household rather than for himself. For wages of $9,000 or more, the behavior of response status is similar to that seen for the middle-range intervals; however, the expected trend for the roundedness of CPS wages is reversed. In other words, among persons in the highest wage class, three-way agreement for rounded CPS wages nearly always exceeds that of nonrounded cases. This is true for three out of the four combinations of interview factors shown in figure 5. Such an occurrence is not unexpected, since any individual reporting wages of $9,000 or more to all three sources will have three-way agreement, regardless of the roundedness of the actual amounts reported.

In order to further illustrate the effect of the interview factors on the levels of agreement, it is interesting to look at three-way agreement by one of the variables. Hence, it was decided to examine roundedness of CPS wages, the variable which appears to be most dominant in figure 5. In the next chart (figure 6), the solid line represents the percent distribution for nonrounded CPS wages and the broken line depicts that for rounded amounts.

Figure 6. --Percent of Three-Way Agreement by IRS Wage Class, Separately by Roundedness of CPS Wages



Three-way agreement for nonrounded CPS wages exceeds that of the rounded cases in every dollar size-class except the last. Furthermore, the sawtoothed effect noticed for the overall distribution in figure 4 is much less evident for persons with nonrounded survey wages. On the other hand, the reverse is true for the corresponding rounded cases: jaggedness of the distribution is far more pronounced than for the total study universe.

This, therefore, further supports the contention made earlier that at least one of the interview factors -- roundedness of CPS wages -- may be a useful predictor of survey wage agreement for the cases examined here. In the near future, regression analysis and other statistical techniques will be used to further study numerous variables, including the interview factors discussed in this paper, in an effort to determine their influence on the extent of agreement among wages from various sources.

FOOTNOTES

* The authors would like to thank H. Lock Oh and Penny Johnston for their assistance in preparing this paper. Helpful editorial comments were received from Ben Bridges, Fritz Scheuren, and Bertram Kestenbaum. Thanks are also due to Catherine Murphy and Deborah Dillard for their typing efforts.

1/ At the 1975 ASA meetings in Atlanta, several papers were presented which focused on the conceptual and reporting differences among the linked CPS, IRS, and SSA data sets. Some of the preliminary analyses presented included comparisons between matched CPS and IRS income data for 1972 [1], and similar comparisons for SSA and IRS wage data [2]. For more information on the basic study see [3] which appears elsewhere in these 1976 Proceedings. See also [4].

As mentioned in most of these referenced papers, all interagency data linkages were performed solely by Census Bureau personnel. Neither IRS nor SSA had access to identified records from each other's files. The tables of matched data used for this paper were produced by the Social Security Administration. However, the file used by SSA could not have been used for other than statistical purposes, since it was simply a random sample of unidentified records.

2/ The population total for CPS-IRS comparable nonjoint returns shown here is slightly smaller than the initial estimate given in [5]. Persons interested in full details should consult [6].

3/ Estimates for these groups and for the analysis of the study universe were obtained using sample weights which incorporated some adjustments for nonmatches and mismatches. These (initial raking) weights are described in detail in [7].

4/ Microfilm files are kept at SSA of the actual wage amounts received from each employer; however, these files were far too expensive to examine on a wholesale basis for the 1973

study. Use has been made of the microfilm files, though, on a subsample basis, as is discussed in [8].

5/ In order to take this factor into account, these distributions were examined by both $500 and $600 intervals. The pattern which emerged was similar for both cases.

6/ Copies of the tables from which these distributions were taken were available in handouts distributed after the session and can be obtained from the authors by writing to them at--Division of Economic and Long-Range Studies, Office of Research and Statistics, Social Security Administration, 1875 Connecticut Avenue, N.W., Washington, D.C. 20009.

7/ Two more papers which may be of interest, dealing with other factors affecting CPS reporting of income data appear in these Proceedings. See [3] and [10].

REFERENCES

[1] Herriot, R. and Spiers, E., "Measuring the Important on Income Statistics of Reporting Differences between the Current Population Survey and Administrative Sources," 1975 American Statistical Association Proceedings, Social Statistics Section, pp. 147-158.

[2] Millea, M.T. and Kilss, B., "Exploration of Differences between Linked Social Security and Internal Revenue Service Wage Data for 1972, "1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 138-146.

[3] Vaughan, D., and Yuskavage, R., "Investigating Discrepancies between Social Security Administration and Current Population Survey Benefit Data for 1972," 1976 American Statistical Association Proceedings, Social Statistics Section.

[4] Studies from Interagency Data Linkages, Report No. 4, Social Security Administration, 1975.

[5] Sailer, P., and Vogel, L. "Exploration of Differences between Linked Current Population Survey and Internal Revenue Service Income Data for 1972," 1975 American Statistical Association Proceedings, Social Statistics Section, 1976, pp. 129-137.

[6] Studies from Interagency Data Linkages, Report No. 9, Social Security Administration, (in preparation).

[7] Studies from Interagency Data Linkages, Report No. 8, Social Security Administration, (in preparation).

[8] Kestenbaum, B., "Evaluating SSA's Current Procedures for Estimating Untaxed Wages," 1976 American Statistical Association Proceedings, Social Statistics Section.

[9] Knott, J.J., "An Analysis of the Effect of Income Rounding in the Current Population Survey," 1971 American Statistical Association Proceedings, Social Statistics Section.

[10] Stevens, J. and Bailar, B., "The Relationship between Various Interviewer Characteristics and the Collection of Income Data," 1976 American Statistical Association Proceedings, Social Statistics Section.