QUALITY OF OCCUPATION INFORMATION ON TAX RETURNS

5.05

Gloria S. Koteen and Paul E. Grayson, Internal Revenue Service

This paper reports on the feasibility and quality of coding occupation information from individual income tax returns. It is part of the collaborative effort of the National Center for Health Statistics (NCHS), Internal Revenue Service (IRS), Social Security Administration (SSA), and the Census Bureau to create linked statistical samples for mortality research. Since there is much interest in exploring the possible relationship between occupation and mortality, it is important to examine the quality of the occupation information available on the tax return to determine if it can be used as a reliable source in this research.

Organizationally, the paper is divided into five parts. Section 1 provides an overview of the most recent (Tax Year 1973) IRS study of the occupation information on individual income tax returns. In Section 2, the 1973 results are provided on "classifiability," i.e., the extent to which returns or taxpayers could be classified by occupation. Comparisons are then made between IRS and the Census Bureau's Current Population Survey (CPS) data in Section 3. Section 4 summarizes the findings of earlier related IRS efforts, and the concluding section of the paper describes the further work on IRS occupation coding currently underway.

1. OVERVIEW OF TAX YEAR 1973 STUDY

The most recent and comprehensive study of IRS occupation information was conducted in 1975, using the 1973 Taxpayer Usage Sample [1]. The Taxpayer Usage Sample is selected each year to examine the changing reporting characteristics of taxpayers. The sample generally represents the majority of all taxpayers in any given year. For Tax Year 1973, it consisted of a systematic sample of returns received at all IRS service centers from January 1, 1974, through May 9, 1974, accounting for about 96 percent of all returns (Forms 1040 and 1040A) filed for Tax Year 1973. The specified sampling rate was 1:13,000.

The study sample consisted of 6,158 returns; 3,294 or 53.5 percent were joint returns (i.e., returns of married couples filing jointly). The 2,864 remaining returns were filed nonjointly. These include single taxpayers, married persons filing separately, unmarried heads of households, and qualifying widow(er)s with dependent children. It was for all these sample cases that the occupation entry on the return was examined.

Since IRS does not do this routinely, the taxpayer entry for "occupation" had to be specially abstracted from each sample return for the study. Classification of taxpayer occupation was then carried out by one of two methods, the "direct" or "indirect." The direct method was based solely on the taxpayer entry in the occupation box; the indirect method was based on other information available on the return, such as adjusted gross income (AGI), wages and type of employer (from Form W-2), and business activity (from Schedules C & F of Form 1040, reporting nonfarm and farm proprietorships, respectively). The indirect method was used when the occupation entry was missing, illegible or too broad to classify, such as "government worker." An application of the indirect method was a case of "no entry," with the Form W-2 listing the name of an engineering firm and showing wages of $30,000. The taxpayer was classified as a "professional, technical or kindred worker," one of the broad occupational classes used in the 1970 Census.

Since the term "occupation" in our study was used broadly to indicate the major activity of the taxpayer, two types of categories were used for coding. For those in the employed civilian labor force, we applied the twelve major classes that were used in the 1970 Census. For those <u>not</u> in the employed civilian labor force, five additional "occupational" classes were defined: housewives, students, retirees, Armed Forces members, and unemployed or disabled workers. (The Technical Note at the end of the paper provides more information on all the occupation classes.)

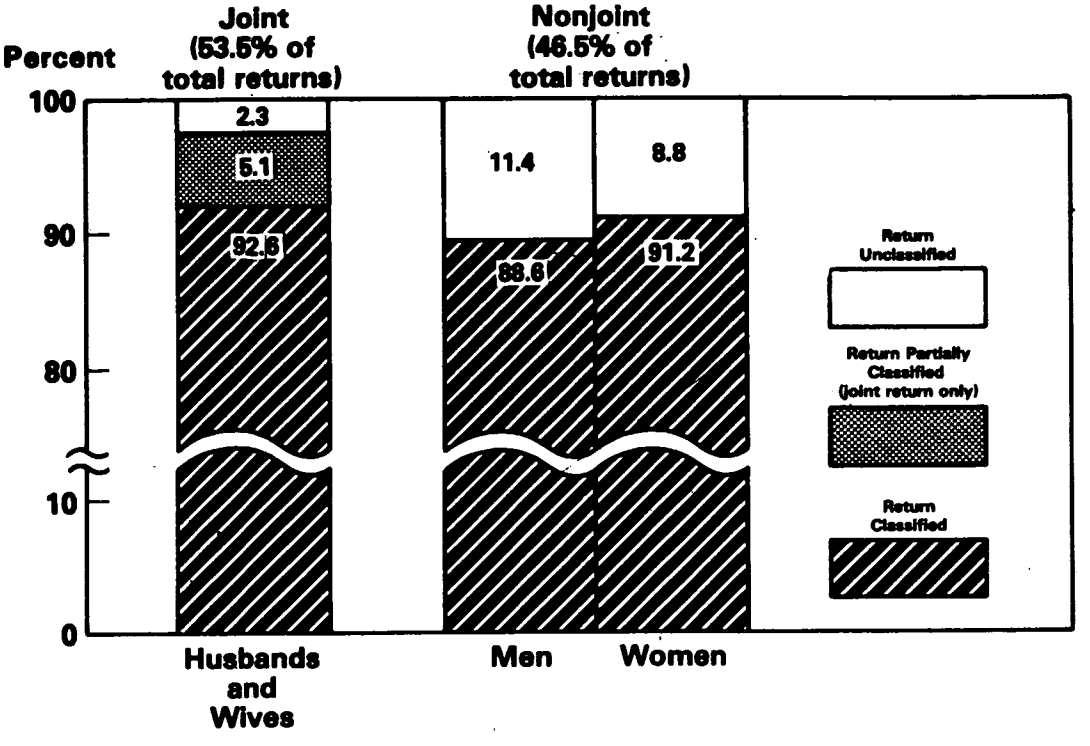
2. CLASSIFIABILITY RESULTS

Considering the entire sample, 91.3 percent of all returns (joint and nonjoint) could be classified by occupation. As shown in Figure 1 for nonjoint returns, 88.6 percent of the men and 91.2 percent of the women could be classified. Among joint returns, occupations of both husband and wife could be classified on 92.6 percent. The classifiable proportion was further subdivided as seen in Figure 2: 83.7 percent of the joint returns could be classified solely from entries made by both taxpayers in the occupation box of the return (direct method), and an additional 8.9 percent could be inferred from other information present on the return (indirect method). An additional 5.1 percent of joint returns could be partially classified (i.e., for one taxpayer only), as shown in Figure 1. For this group, the direct method was about as important as the indirect method (Figure 2). Only 2.3 percent of all joint returns thus remained with neither taxpayer classified.

When comparing the classifiability of individuals on joint versus nonjoint returns as in Figure 3, it is evident that husbands and wives on joint returns had a higher level of total classifiability--93.5 percent and 96.8 percent respectively--than men and women on nonjoint returns, and a higher direct classification percentage as well. This is probably due to the fact that joint returns had about 4 percent more completed occupation entries than nonjoint

Figure 1

Occupational Classifiability of Returns by Filing Status and Sex (nonjoint returns), Tax Year 1973



Question: **What Percentage of Returns Could be Classified?** (91.3 percent of *all* returns (joint & nonjoint) could be classified by occupation)

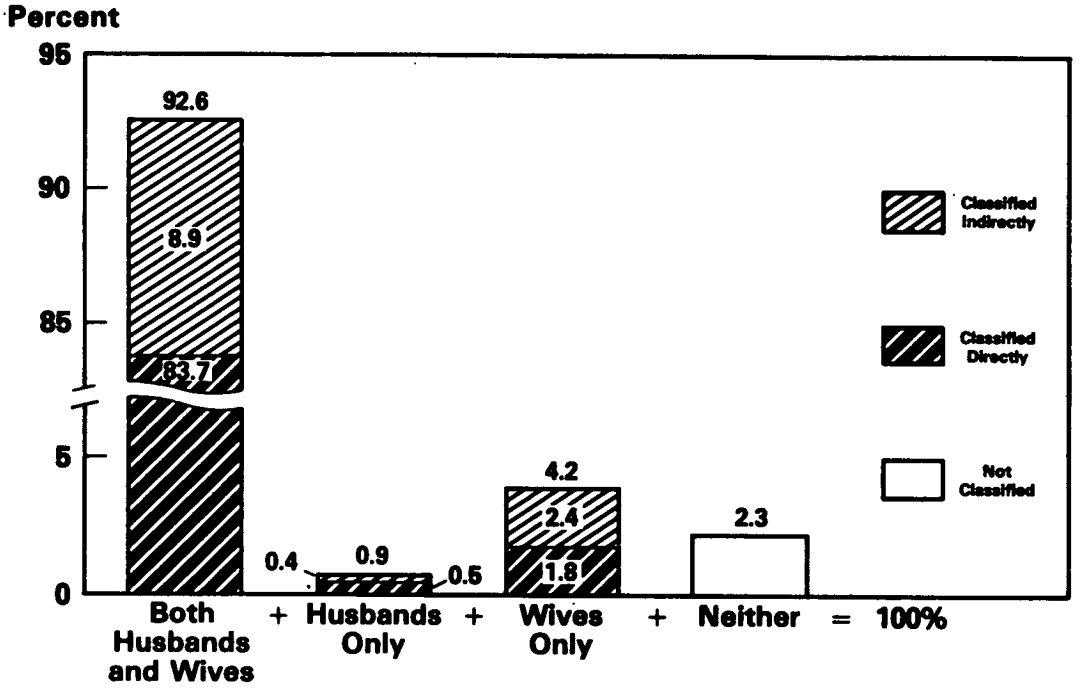
a. *On joint returns*
   1. Occupations of both husbands and wives could be classified on 92.6 percent of the joint returns.
   2. An additional 5.1 percent of joint returns could be partially classified (for one taxpayer only on the joint return)

b. *On nonjoint returns*
   1. 88.6 percent of the men and 91.2 percent of the women could be classified.

Source, IRS data: Based on the 1973 Taxpayer Usage Sample — a systematic sample of 6100 returns received at all IRS service centers January 1 — May 9, 1974, accounting for about 98 percent of all returns (Forms 1040 and 1040 A) filed for Tax Year 1973 during 1974. (The specified sampling rate was 1:13,000).

Figure 2

Distribution of Joint Returns by Classifiability of Taxpayer's Occupation and Method of Classification, Tax Year 1973

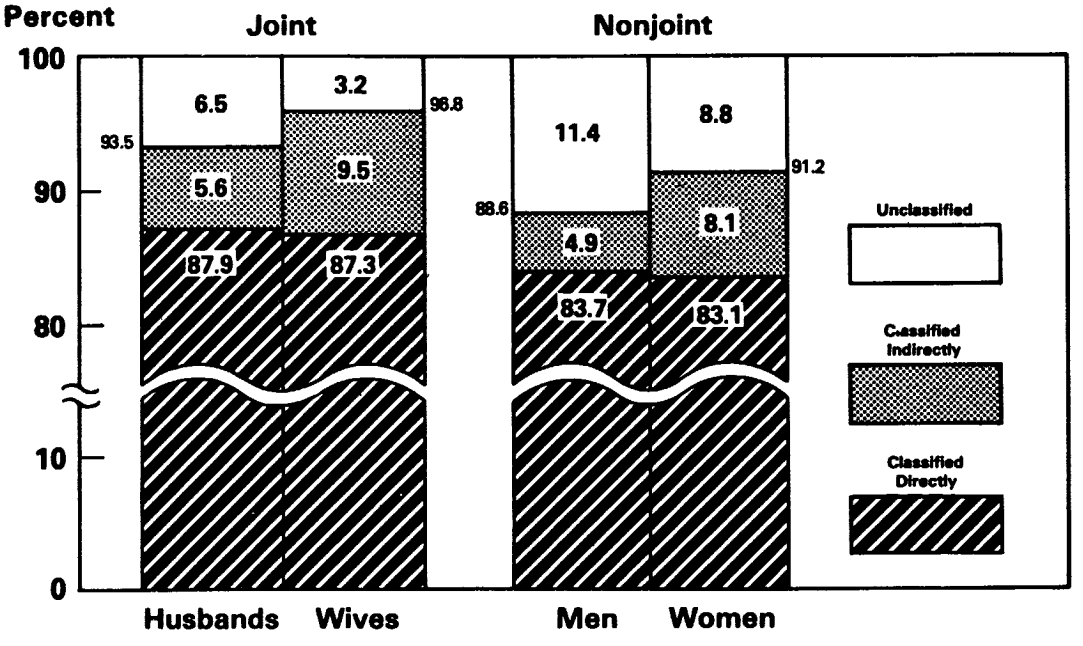


Question: **How are Joint Returns Distributed with Regard to Classifiability of Occupation?**

1. 83.7 percent of joint returns could be classified directly (from entry in occupation box).
2. An additional 8.9 percent could be classified indirectly, from other information on the return
3. The 5.1 percent partially classified (see Figure 1) consists of 0.9 + 4.2, where only one taxpayer, but *not both* on a joint return, could be classified. (For this group, the direct and indirect method were equally important.)
4. Only 2.3 percent of joint returns were unclassifiable.

Figure 3

Occupational Classifiability of Individuals Filing Tax Returns, By Filing Status, Sex and Method of Classification, Tax Year 1973



Question 1: **Could Relatively More Husbands and Wives on Joint Returns be Classified by Occupation than Men and Women on Nonjoint Returns?** ***YES,*** because 4 percent more husbands on joint returns filled in the occupation entry box than did men on nonjoint returns. The same differential was found for wives on joint returns compared with women on nonjoint returns.

Question 2: **Could Relatively More Returns (Both Joint and Nonjoint) Filed by Women be Classified by Occupation than Returns Filed by Men?** ***YES,*** because more returns filed by women could be classified *indirectly.* The differential was about the same, irrespective of filing status.

returns. The evidence suggests that taxpayers filing joint returns may tend to check and reinforce one another.

When men and women were compared (Figure 3), a larger percentage of female occupations than male could be classified indirectly, resulting in a higher total percentage of women classified than men. This difference was equally true for both joint and nonjoint filing status. No doubt, a contributing factor was the frequent use of the "housewife" classification whenever a woman's Form W-2 was missing and no evidence of self-employment existed.

3. STUDY RESULTS COMPARED WITH CURRENT POPULATION SURVEY

To test the reasonableness of our classifications, we compared the study results with the occupational distribution for 1973 published as a result of the March 1974 Current Population Survey (CPS) [2]. For the comparison, we could use only the 80.1 percent of the IRS males and 49.3 percent of the IRS females classified in the employed civilian labor force (Table 1). The occupational distribution for this portion of the sample is compared with the CPS in Table 2.

Table 2 shows that the results of our IRS coding closely parallel the CPS distribution, both in terms of level and rank. Several examples of close agreement are males in "sales workers" and "clerical" occupations, and males and females in "operatives, except transport." Levels among women did not compare as closely as among men. However, for both men and women, the ranking of occupations from the highest to the lowest percentage for IRS and CPS were very similar. In fact, the Spearman rank coefficient yielded a value of 0.97 for males and 0.89 for females, where 1.0 would be perfect correspondence.

It is interesting to note that three of the differences, where IRS percentages are lower than CPS', are for low wage occupations where total income is often below the level required to file a return. This can be seen for female "private household workers" and for male and female "farm laborers and foremen." In other cases, where IRS percentages are higher than CPS', as for "professionals" and "managers," research by SSA has indicated that when adjustments are made to exclude nonfilers in the CPS, many of these gaps between IRS and CPS are closed [3]. Even after dividing men and women into marital status groups, we continued to find parallel distributions between IRS and CPS percentages (Table 3).

4. PREVIOUS IRS STUDIES INVOLVING OCCUPATIONAL CODING

During the past 15 years, several other smaller scale studies were conducted at IRS, or with IRS cooperation, involving occupational coding from tax returns. These included the 1963 Pilot Link Study conducted by SSA and two studies for 1968 and 1970 conducted at IRS. All of these studies found it feasible to code tax returns into broad occupational classes.

1963 Pilot Link Study [5].--SSA coded occupation entries abstracted from 1963 tax returns for persons identified in the CPS. When comparing the occupation codes of IRS with CPS for approximately 3,400 matched cases, the occupations agreed in nearly three-fourths of the cases. However, only 60 percent of the abstract sheets had known or codable IRS occupations, probably due to errors in clerical processing or to an abstract sheet whose design made it easy to miss the occupation entry. (See Table 4.)

Tax Year 1968 Study [6].--For a sample of 600 men, 88 percent of the taxpayers could be classified in labor force classes, another 6 percent in non-labor force classes and about 6 percent of the taxpayers failed to indicate any occupation. On joint returns, only the husband's occupation was used. The classification system was the same as that used in the 1970 Census.

Tax Year 1970 Study [7].--This study again was based on a sample of about 600 returns, taken from only one IRS region. On joint returns, only the primary taxpayer was used. A significant feature of this study was coding on the basis of major source of income. The taxpayer was classified "retiree" when pensions or annuities were the major source; he was classified "investor" when dividends, interest or capital gains were the major source. In the study, use was also made of "Dunn and Bradstreet" to provide insights as to industry so that a better code for occupation could be obtained.

5. CONCLUSIONS AND AREAS FOR FURTHER RESEARCH

The study results presented in this paper from Tax Years 1963, 1968, 1970, and especially from the large 1973 study, strongly suggest that the tax return can be used as a reliable source for occupation information, at least if broad occupational classes are employed. Classifiability by occupation was greater than 90 percent, and the IRS occupation distribution closely mirrored the CPS, especially after adjusting CPS figures for nonfiling.

With the interest in occupation data to meet mortality and other research needs, and encouraged by the results of the studies to date, IRS and SSA plan to embark on yet another effort of this sort; proposals are currently under consideration to code occupation for all of the individual income taxpayers in the 1979 Statistics of Income (SOI) sample. The SOI sample for 1979 will consist of nearly 170,000 returns with over a quarter of a million taxpayers.

In an effort to construct detailed occupation codes for this new project [8], two major sources of information will be tapped: the new alphabetical Standard Occupational Classification (SOC) Manual will be used to help code occupations from the tax returns, and data from SSA's file of employers will be introduced to provide

supplemental information on industry. Then, each occupational title on the tax return can be sorted by industry, providing an industrial context for each code. This, no doubt, will result in more effective classification overall.

The Tax Year 1976 Taxpayer Usage Sample is currently being coded employing the operation just described. (The 1976 sample consists of about 6,900 returns.) From this further preliminary work, IRS hopes to learn several things, namely: (1) how well occupation entries on tax returns conform to the SOC, (2) how many entries are codable, (3) whether the industry codes are helpful and how best to use them, and (4) what level of detail (of the four-level coding system) can be achieved. IRS also hopes to work out any technical problems that may arise from keypunching occupation entries and employer identification numbers and matching them to the SSA files.

Eventually, these efforts may lead to IRS' creation of its own classification reference tape to be used as a basis for coding occupations from the 1979 SOI file. Such a tape would consist of an alphabetical listing of each occupation-industry combination, with an SOC code assigned to each. Any occupation not covered by this tape will be dealt with individually, in the same manner used to create the tape; that is, sorting by industry code within each occupational title and coding according to the SOC Manual. Each new coded occupation-industry combination will then be added to the tape to increase the known elements and further reduce the number of occupations needing individual attention.

Of course, classifiability--even improved by supplemental industry data--is only "part of the game." An occupation code is not very useful if it is incorrect. Hence, a certain amount of validation work is also necessary. Efforts on this behalf are currently being explored with the Census Bureau. Under consideration at this time is a proposal that would require the Bureau to match approximately 70,000 individuals from the 1979 SOI file to the 1980 Census on a name and address basis. Since occupation and industry data will be available for about 20 percent of these cases, the occupation codes constructed in SOI could be compared to those in the decennial census for approximately 14,000 individuals.

TECHNICAL NOTE

MAJOR OCCUPATIONAL CLASSES FOR TAX YEAR 1973 STUDY, WITH EXAMPLES BASED ON THE CENSUS CLASSIFICATION SYSTEM

Employed Civilian Labor Force

Professional, Technical, and Kindred workers
Managers and Administrators, except farm (includes all store owners except barbers and beauticians)
Sales Workers (includes self-employed Avon saleslady, for example)
Clerical and Kindred Workers (includes checkout clerks)
Craftsmen and Kindred Workers (includes mechanics)
Operatives, except transport (includes factory and production workers such as textile workers with entry of "Textiles" in occupation box)
Transport Equipment Operatives
Laborers, except farm (includes all unskilled construction workers such as those with entry of "Construction" in occupation box)
Private Household Workers
Service Workers, except private household (includes self-employed barbers and beauticians)
Farmers and Farm Managers
Farm Laborers and Foremen

Other "Occupations"

Housewives (with and without earnings)
Retirees (with and without earnings)
Students (with and without earnings)
Uniformed members of the Armed Services
Unemployed or Disabled workers (considered unemployed if no earnings are recorded even though an occupation, other than housewife, student, or retired, is entered)

ACKNOWLEDGEMENTS

The authors would like to thank several SSA staff members, especially Wendy Alvey and Beth Kilss, for their help in presenting this paper at the meetings. At IRS the authors also owe a debt of gratitude to several people for helpful advice, especially to Pete Sailer and Janet Barnhardt. Typing assistance was provided by Vivian Louallen at IRS and Joan Reynolds at SSA.

REFERENCES

[1] Koteen, G., "Occupations Reported on Individual Tax Returns-- Tax Year 1973," memorandum dated August 28, 1975, Statistics Division, Internal Revenue Service. See LASS Working Notes No. 2, Office of Research and Statistics, Social Security Administration, January 30, 1979, pp. 1-13.

[2] Work Experience of the Population, 1973, Special Labor Force Report 171, U.S. Department of Labor, BLS, 1975. Table A-4. "Occupation: Persons with work experience in 1973, by longest job and sex," p. A-13.

[3] Johnston, M.P. "Occupations of CPS Taxfilers in 1973 Simulated Tax Units," Unpublished SSA working note, 1979. See also, Sailer, P.J. and Vogel, L. "Exploration of Differences between Current Population Survey and Internal Revenue Service Income Data for 1972," American Statistical Association 1975 Proceedings, Social Statistics Section, pp. 129-137.

[4] Marital and Family Characteristics of the Labor Force in March 1973, Special Labor Force Report 164, U.S. Department of Labor, BLS, 1974. Table E. "Major occupation group of employed persons, by full time and part-time status, marital status, sex, and selected age groups, March 1973," p. 17. (This table includes only those persons employed in March 1973, unlike the table in reference [2] above which is for all persons who worked at any time during the year 1973.

[5] Aziz, F., Kilss, B. and Scheuren, F., "Occupational Coding from Tax Returns in the Pilot Link Study -- Tax Year 1963." See LASS Working Notes No. 2, pp. 27-58.

[6] Reiser, B.S., "Occupation Data Reported on Individual Income Tax Returns -- Tax Year 1968," memorandum dated March 12, 1970, Statistics Division, Internal Revenue Service. See LASS Working Notes No. 2, pp. 21-26.

[7] Sailer, P.J. and Robinson, C. "Feasibility of Occupational Coding from Tax Returns-- Tax Year 1970," memorandum dated July 29, 1971, Statistics Division, Internal Revenue Service. See LASS Working Notes No. 2, pp. 14-20.

[8] Sailer, P.J., "Final Proposed Procedure and Related Cost Estimates for Occupational Coding of SOI Data," December 27, 1978. See LASS Working Notes No. 3, January 30, 1979, pp. 170-173.

Table 1.--Percentage Distribution of Occupations for Individuals with Classified Entry on Sample Returns: by Sex and Filing Status, Tax Year 1973

Occupation Classes	Male			Female		
	Total	Joint	Non-joint	Total	Joint	Non-joint
Grand total	4806	3294	1512	4646	3294	1352
Number unclassifiable	384	212	172	225	106	119
Number classifiable	4422	3082	1340	4421	3188	1233
Total, percent	100.0	100.0	100.0	100.0	100.0	100.0
EMPLOYED CIVILIAN LABOR FORCE						
Total	80.1	85.5	67.9	49.3	41.1	70.4
Professional, technical	13.0	15.1	8.4	9.0	8.1	11.5
Managers and administrators	11.2	14.1	4.5	2.6	2.0	4.5
Sales workers	5.3	5.8	4.0	2.9	2.5	3.7
Clerical	5.3	4.5	7.1	18.8	15.3	27.7
Craftsmen	14.9	17.0	10.3	0.4	0.3	0.4
Operatives, except transport	9.6	9.8	9.3	6.2	5.5	7.9
Transport equipment operatives	4.1	4.5	3.1	0.2	0.3	0.1
Laborers, except farm	8.1	5.9	13.1	1.2	1.0	1.9
Private household workers	-	-	-	0.5	0.3	0.7
Service workers	5.2	4.8	6.0	7.4	5.8	11.8
Farmers and farm managers	2.8	3.4	1.3	0.1	(*)	0.2
Farm laborers and foremen	0.6	0.6	0.8	(*)	(*)	-
OTHER "OCCUPATIONS"						
Total	19.9	14.5	32.1	50.7	58.9	29.3
Housewives	-	-	-	40.7	55.8	1.8
Retirees	7.6	8.9	4.8	4.3	2.5	8.6
Students	6.7	1.1	19.5	5.2	0.5	17.4
Uniformed armed services	4.7	3.7	6.8	0.1	-	0.4
Unemployed or disabled	0.9	0.8	1.0	0.4	0.1	1.1

Note: Detail may not add to totals due to rounding; (*) denotes less than 0.05 percent.

Table 2.--Occupational Distribution of Employed Civilian Labor Force by Sex: IRS Study Results for Tax Year 1973 Compared with March 1974 CPS Information for Calendar Year 1973

Major Occupation Classes	Male		Female	
	IRS	CPS	IRS	CPS
Total, percent.....	100.0	100.0	100.0	100.0
Professional, technical.....	16.3	13.0	18.3	14.0
Managers and administrators..	14.0	12.9	5.5	4.5
Sales workers.....	6.6	5.8	5.8	7.3
Clerical.....	6.6	6.3	38.1	33.1
Craftsmen.....	18.7	20.7	0.7	1.7
Operatives, except transport.	12.0	12.7	12.5	12.7
Transport equipment workers..	5.1	5.8	0.4	0.5
Laborers.....	10.0	8.9	2.5	1.0
Private household workers....	0.1	(*)	0.9	4.1
Service workers.....	6.4	8.6	15.1	18.6
Farmers and farm managers....	3.4	2.9	0.2	0.5
Farm laborers and foremen....	0.8	2.4	(*)	2.1

Note: IRS study percentages based on adjusted totals as explained in text. CPS data from [2]. Detail may not add to totals due to rounding; (*) denotes less than 0.05 percent.

Table 3.--Occupational Distribution of Employed Civilian Labor Force, by Sex and Marital Status: IRS Tax Year 1973 Study Results Compared with the March 1973 CPS

Major Occupational Classes	Male				Female			
	"Married"		Other than "married"		"Married"		Other than "Married"	
	IRS	CPS	IRS	CPS	IRS	CPS	IRS	CPS
Total, percent.....	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Professional, technical....	12.3	11.2	17.6	14.5	16.3	13.6	19.6	16.1
Managers and administrators	6.6	6.4	16.5	16.0	6.3	4.7	4.9	5.2
Sales workers.....	5.9	5.4	6.8	6.5	5.3	6.1	6.2	7.2
Clerical.....	10.4	9.1	5.3	5.9	39.3	34.2	37.2	34.1
Craftsmen.....	15.2	14.4	19.9	23.1	0.6	1.2	0.8	1.5
Operatives except transport	13.6	15.4	11.4	12.0	11.3	11.5	13.3	14.5
Transport equipment workers	4.6	5.7	5.3	6.2	0.1	0.3	0.6	0.6
Laborers.....	19.2	13.3	6.9	5.0	2.8	1.0	2.4	0.7
Private household workers..	0.1	0.2	-	(*)	1.0	6.8	0.8	2.6
Service workers.....	8.8	14.1	5.6	6.2	16.7	19.6	14.0	16.0
Farmers and farm managers..	2.0	1.7	4.0	3.4	0.3	0.3	0.1	0.3
Farm laborers and foremen..	1.2	3.2	0.7	1.1	-	0.7	0.1	1.3

Note: IRS Study percentages for married persons were obtained by treating as "married" only joint taxpayers. This was done in an attempt to roughly approximate the CPS "married" data (from [4]) which was based only on married persons living with their spouses. Note also that detail may not add to totals due to rounding; (*) denotes less than 0.05 percent.

Table 4.--1963 Pilot Link Study: IRS and CPS Occupational Classifications Compared

IRS Classification	CPS Classification									
	Total	Professional technical and kindred workers	Managers and administrators except farm	Clerical and kindred workers	Sales workers	Craftsmen and kindred workers	Operatives including transport equipment operatives	Service workers except private household	Laborers except farm	Diagonal total
Total.....	3,362	598	438	615	190	533	561	325	102	3,362
Professional, technical, and kindred workers.....	652	529	32	27	8	19	13	24	0	502
Managers and administrators, except farm.....	396	24	259	30	12	28	14	26	3	0
Clerical and kindred workers..	624	24	31	502	34	9	10	8	6	27
Sales workers.....	204	3	40	15	126	3	13	4	0	45
Craftsmen and kindred workers.	638	8	51	6	4	419	130	2	18	41
Operatives, including transport equipment operatives....	441	2	11	22	2	35	345	9	15	68
Service workers, except private household.....	290	6	11	6	3	4	11	244	5	78
Laborers, except farm.....	117	2	3	7	1	16	25	8	55	243
Diagonal total.....	3,362	381	2	9	20	26	95	101	128	2,479

Source: Derived from the 1963 Pilot Link File documented in Report No. 7 in the Studies from Interagency Data Linkages series, Social Security Administration.

Note: The diagonal totals were obtained by adding up the table entries in a diagonal direction. Consider for example, the main diagonal, which has been underlined. If one sums these underlined figures, the diagonal total obtained is shown in the lower right-hand corner of the table. Also shown are diagonal totals for cells above and below the main diagonal. For more details on tabular displays such as this, see Scheuren, F. J. and Oh, H.L., Communications in Statistics, July 1975.