

1979 OCCUPATION CODING STUDY/1979-1983 MORTALITY STUDY

Bobby Clark, Dodie Riley and Peter Sailer, Internal Revenue Service
Internal Revenue Service (R:S:I), 1111 Constitution Ave., N.W., Washington, DC 20224

TAXPAYER DATA CLASSIFIED BY OCCUPATION

The purpose of this paper is to discuss the results of an occupation-coding effort which has been ongoing at the Internal Revenue Service (IRS), and to show how it was used to study occupation-related mortality issues. A previous paper presented by members of the IRS' Statistics of Income Division (SOI) on coding taxpayer entries on occupation stated that the IRS would produce occupation data [1]. It is, therefore, a pleasure for us to announce that, indeed, we have produced occupation data. Moreover, the occupation code is but one of many enhancements to our 1979 Statistics of Income File. It also has indicators of the sex of the taxpayer, and, when a joint return was filed, amounts of income from several sources are broken out separately for the husband and wife. Industry codes were obtained for both sole proprietors and for employees, and cause of death codes were added for taxpayers who died between 1979 and 1983.

Users of SOI data should be aware that attempts to divide up the salaries of joint return filers between those of the husband and the wife have been made only twice previously--for Tax Years 1969 and 1974. And there has been only one previous attempt to code the little box on the Form 1040 returns that says: "Your Occupation". That was for Tax Year 1916, the first year of the Statistics of Income series [2]. In preparing the 1916 report, our predecessors had to settle for some rather broad occupational categories, including one called "Corporation officials, secretaries, managers, cashiers, presidents, etc." Adding cause of death to a file of 1040's was a first time effort.

Organizationally, this paper is divided into five parts. First, we present a very brief review of the methodology used to occupation-code the SOI file for Tax Year 1979, and compare some of our tabulations to occupation data available from the decennial Census. (The Census Bureau very kindly decided to hold a census covering the same income year for which we prepared our tax return statistics [3].) Next, we discuss our 1979-1983 Mortality Study conducted primarily for the National Cancer Institute (NCI) and how the two files were linked to produce some occupation-related mortality statistics. Finally, we discuss plans for the future and some preliminary conclusions.

COMPARISON OF CENSUS AND IRS OCCUPATION-CODING METHODOLOGIES

Before comparing the Census and IRS statistics, it should be pointed out that the documents from which the two agencies obtained their data--and the methodologies used to code them--differed substantially. The most obvious difference is the fact that the tax returns contain a little box asking for "Your Occupa-

tion," whereas the Census Bureau was able to include two questions on its questionnaire: "What kind of work were you doing?" and "What were your most important activities or duties?"

Both systems used the industry code in some instances to generate an occupation code. Here we may have had a slight advantage over Census --at least, our method was less work. Through a series of matches from one administrative records file to another, we were able to access the Social Security Administration's (SSA) file of all employers in the U.S., which contains a Standard Industrial Classification (SIC) Code for each employer; in the case of multi-unit firms, it even contains, in many cases, separate codes for each unit [4]. At Census, on the other hand, coders had to code the answers each respondent gave to a three-part question on the nature of the firm for which they worked [5].

The Census form contains instructions to the respondent which indicate the level of detail being sought (for example, "nurse" is unacceptable, say "registered nurse"). The total IRS instruction was: "Remember to show your occupations in the spaces in the upper right corner just below the social security number blocks."

Census used a modified form of the Standard Occupational Classification System to code its questionnaires [6]. In the end, we had to modify the SOC slightly for our use as well, but not to the same extent nor always in the same way.

In addition to the Standard Occupational Classification (SOC) Manual [7], we used coding aids prepared by the Labor Department, notably their Dictionary of Occupational Titles [8]; Census prepared its own coding aids.

The most important difference, however, is that, at the time of the 1980 Census, occupation coding was still a manual operation at the Census Bureau. (They were, of course, working on a computerized occupation and industry coding scheme of their own [9]). With the help of our computerized occupation coding dictionary, we feel that we achieved a high degree of consistency in our coding process. Some of the decisions we incorporated into the dictionary may have been wrong, but at least they were consistently wrong. After several years of analyzing and reviewing our entries, we feel that, by and large, we are consistently right (within statistical tolerances).

A Comparison of Census and IRS Occupation Distributions

In this section, we will compare occupation distributions from the Census Bureau's 1980 Decennial Census of Population to those tabulated from our 1979 SOI file. The Census data are for 104 million individuals in the Civilian Labor Force [10], which includes anybody working at the time of the Census, unpaid family workers, as well as unemployed individuals with work experience between 1975 and 1980. Our data are

for a little over 98 million tax return filers who indicated labor force occupation titles (other military) on their tax returns and had earned income for Tax Year 1979. The major difference, of course, is individuals who were not required to file because they did not meet the minimum filing requirements [11]. Because of the conceptual differences between the IRS and Census data, we had some special tabulations prepared for the Census public-use microdata files from which we excluded all those individuals who fell below the filing requirement. The adjusted Census number was just under 89 million. Some of the 9 million who appear in our distribution and not on the Census adjusted distribution are individuals who were not required to file but filed anyway to receive a refund or claim the earned income credit.

Three-way comparisons among our data, Census published data, and the special tabulations from the Census public-use microdata sample [12] are shown in Table 1.

The distributions in Table 1 compare our figures for each two-digit SOC code with those from Census [13]. For the sake of completeness, two sets of Census figures are shown: Those published by the Census Bureau for the Civilian Labor Force, and those produced from the Census public-use file which exclude individuals with income below the filing requirements. The IRS figures are for taxpayers with earned incomes in Tax Year 1979, excluding those with non-civilian labor force occupation entries (such as "Military" or "Student").

There are some occupations where the adjusted Census figures are so close to those from IRS as to be within expected sampling variability. These include, for example, judges and lawyers, as well as realtors, school teachers, farmers, and administrative support occupations (such as secretaries and other clerical personnel). All in all, 60 percent of all taxpayers worked in occupations where the IRS total was within 10 percent of the Census Bureau total, and an additional 28 percent worked in occupations where the IRS was either between the original and the adjusted Census figures, or within 10 percent of the original figures.

As we were creating our computerized occupation-coding dictionary, it became very evident that many nurses, whether they were registered nurses or licensed practical nurses, used the term "nurse" in their occupation boxes. The SOC manual shows registered nurses as a separate major group, whereas licensed practical nurses are considered health technologists. For the purpose of our statistics, we had to create a combined category which covered both groups.

It was necessary to redefine, for the purpose of tabulating tax return data, some of the occupational categories in the Standard Occupational Classification system. For example, it was very difficult to distinguish between handlers/laborers and material movers. Also, further research showed that many individuals who call themselves engineers on their tax returns are really "engineering technicians" (SOC major group 37), which enabled us to combine these occupations.

There were a few cases where the IRS figure fell significantly short of the Census figure.

All of those groups which heavily identified with government workers are understated, including "officials and administrators" and "social workers." They are presumably included in with the "government workers not elsewhere classified."

One more point needs to be made on the reasons for the differences between the IRS and the Census figures, and that has to do with timing. Nearly one-half of the taxpayers in a small pilot-study who reported different occupations to IRS and to Census also reported different employers. The Census Bureau asked for the occupation during the last week in March, 1980 (the week before the census). Presumably, most taxpayers reported to IRS the major occupation they held in the course of Tax Year 1979. While job changes to and from a given job category should even out in most cases, they could make a large difference in the category "officials and administrators, public administration," which includes state, county, and city-level legislators who may have other jobs during a large portion of the year.

In summary, then, the conclusion of the occupation study is as follows: (1) We concede that it is a bit late for 1979 occupation data; on the other hand, this is the only study currently available that relates occupation to taxes; (2) we proved that it can be done, opening the door to all kinds of future applications; (3) the study provided a basis for doing future studies much more cheaply with the use of an electronic occupation-coded dictionary; and (4) it allows a whole new area of projecting mortality rates by occupation based on administrative records.

1979-1983 MORTALITY STUDY

As mentioned earlier, the main purpose of the 1979-1983 Mortality Study was to test the adequacy of a combination of IRS information on occupation of the deceased as reported on their tax return, and the Social Security Administration's information on industry in which individuals were employed, and to detect differentials in mortality associated with employment. A secondary purpose was to test the feasibility of developing a longitudinal file containing employment histories. In order to obtain retrospective data for this file, a portion of the sample used for this project overlapped with SSA's Continuous Work History Sample (CWHHS). A third purpose of this study was to test the usefulness of the National Death Index (NDI).

The basis of the 1979-1983 Mortality Study was the occupation-coded 1979 Statistics of Income sample. This file consisted of a sample containing 203,536 tax returns representing 358,262 taxpayers (joint returns always represent two taxpayers, even if only one of them has income). The file created was used to match against the NDI to obtain the Death Certificate Numbers. As noted previously, a portion of the sample (about 45,000 returns) was chosen to overlap with the CWHHS. The purpose of this overlap was to give us the ability to obtain longitudinal data from the Social Security Administration which would, for earlier years, at least provide information on the industry in

Table 1.—Comparison of IRS and Census Occupation Data

1980 SOC Occupation Codes (Two-digit) and Titles	Census	Census Adjusted	IRS
Total.....	103,718,076	88,858,400	97,696,202
EXECUTIVE, ADMINISTRATIVE AND MANAGERIAL OCCUPATIONS			
Officials and Administrators, Public Administration (11).....	368,666	353,700	188,546
Officials and Administrators, Other (12-13).....	7,392,520	7,049,200	6,733,184
Management Related Occupations (14).....	2,617,332	2,510,200	2,340,591
NATURAL SCIENTISTS AND MATHEMATICIANS			
Computer, Mathematical and Operations Research Occupations (17)....	330,447	324,400	263,970
Natural Scientists (18) 1/.....	314,195	296,800	340,636
SOCIAL SCIENTISTS, SOCIAL AND RELIGIOUS WORKERS AND LAWYERS			
Social Scientists and Urban Planners (19).....	217,858	208,900	119,277
Social, Recreation and Religious Workers (20).....	830,172	765,800	604,268
Lawyers and Judges (21).....	529,679	504,100	530,906
TEACHERS, LIBRARIANS, AND COUNSELORS			
Teachers; College, University and Other Postsecondary Institutes (22).....	637,149	586,000	301,563
Teachers, except Postsecondary Institutions (23).....	3,722,120	3,479,800	3,660,305
Vocational and Educational Counselors (24).....	198,256	163,100	96,242
Librarians, Archivists, and Curators (25).....	202,681	166,100	161,981
HEALTH DIAGNOSING AND TREATING PRACTITIONERS			
Physicians and Dentists (26) 1/.....	558,546	540,800	591,439
Veterinarians (27).....	34,355	29,700	41,364
Other Health Diagnosing and Treating Practitioners (28).....	53,925	51,900	40,883
Pharmacists, Dietitians, Therapists, and Physician's Assistants (30) 1/.....	437,277	409,400	416,806
WRITERS, ARTISTS, ENTERTAINERS AND ATHLETES			
Writers, Artists, Performers, and Related Workers (32).....	902,629	781,700	923,655
Editors, Reporters, Public Relations Specialists and Announcers (33).....	377,854	333,200	253,872
Athletes and Related Workers (34) 1/.....	52,802	35,600	51,776
ENGINEERS, SURVEYORS, ARCHITECTS, TECHNOLOGISTS AND TECHNICIANS			
Engineers and Related Technologists, Technicians Surveyors, and Architects (16,37).....	3,586,973	3,364,200	3,405,467
Nurses, Health Technologists and Technicians (29,36) 2/.....	2,274,646	2,109,700	2,372,019
Science Technologists and Technicians, Technologists and Technicians; except Health and Engineering (38-39).....	1,099,658	990,000	801,846
MARKETING, SALES, ADMINISTRATIVE SUPPORT, AND SERVICE			
Supervisors; Marketing, Sales, Administrative Support, and Service Occupations (40).....	2,660,710	2,517,400	2,520,333
Insurance, Securities, Real Estate, and Business Service Sales Occupations (41).....	1,846,382	1,741,600	1,784,699
Sales Occupations, Commodities except Retail (42).....	1,301,246	1,215,400	1,290,904
Sales Occupations, Retail (43).....	5,499,267	3,856,000	3,263,360
Administrative Support Occupations, Sales, Service, and Related Occupations Including Clerical (44-47).....	16,485,484	14,359,100	14,938,417
Private Household Occupations (50).....	626,945	290,600	19,542
Protective Service Occupations (51).....	1,541,901	1,349,500	1,382,098
Service Occupations, except Private Household and Protective (52)..	11,437,228	7,746,100	7,265,382
AGRICULTURAL, FORESTRY AND FISHING OCCUPATIONS			
Farm Operators and Managers (55).....	1,314,773	1,133,800	1,197,743
Other Agricultural and Related Occupations (56).....	1,503,038	954,700	1,309,328
Forestry and Logging Occupations (57).....	150,435	117,500	95,592
Fishers, Hunters and Trappers (58).....	56,884	42,800	68,079
CONSTRUCTION, EXTRACTIVE, AND PRODUCTION INCLUDING MECHANICS AND REPAIRERS 3/			
Mechanics and Repairers (60-61).....	3,982,525	3,659,200	4,383,581
Construction Trades (631, 641).....	4,814,117	4,319,024	3,677,963
Extractive Occupations (632, 651).....	315,587	289,377	323,845
Precision Production and Production Occupations (67-78).....	14,525,877	13,008,500	10,377,107
TRANSPORTATION, MATERIAL MOVING, AND EQUIPMENT HANDLING AND CLEANING OCCUPATIONS 3/			
Transportation (81-82).....	3,576,038	3,206,933	3,185,022
Handlers, Material Movers, Equipment Cleaners, Laborers, and Helpers (83, 85-87).....	6,413,690	4,963,766	6,229,133
GOVERNMENT EMPLOYEES NOT ELSEWHERE CLASSIFIED	N/A	N/A	1,428,611
NON-CODABLE TAXPAYERS WITH EARNED INCOME 4/	N/A	N/A	8,714,867

1/No imputations allowed in these categories.

2/"Registered Nurses" (29) have not been shown separately in this table. As they could not be coded separately from other nurses, IRS had to combine them with Licensed Practical Nurses (3660). They are, therefore, shown as part of the "Health Technologists and Technicians" (36) group.

3/SOC shows supervisors as separate groups; however, they have been combined with the workers they supervise for this comparison.

4/No occupation or industry data available.

Sources: IRS figures are based on unpublished tabulations of the occupation-coded 1979 Statistics of Income sample. Census figures are taken from "Table 1. Detailed Occupation of the Civilian Labor Force by Sex, Race, and Spanish Origin: 1980," 1980 Census of Population: Detailed Occupation and Years of School Completed by Age, for the Civilian Labor Force by Sex, Race and Spanish Origin: 1980 Supplementary Report, PC80S1-8, Bureau of the Census, March 1983. The Census adjusted figures are based on special tabulations from the 1980 Census public-use microdata sample which exclude individuals with incomes below the filing requirements.

which the taxpayer worked. For future years, it would be possible to add not only the industry data from SSA, but also to pre-identify these taxpayers during IRS statistical processing and obtain occupation data.

During the regular SOI editing of the Tax Year 1979 returns, the editors were asked to enter a sex code for each taxpayer, based on the taxpayer's first name. The sex of the decedent is shown on the Death Certificate.

Commingling Records

In order to be able to order death certificates from the States without disclosing fact of filing, we were required by our Disclosure Division to "commingle" or combine our death certificates with those of other projects that did not involve individual income tax returns. NCI provided most of the records for commingling, with the Foreign Branch of the Statistics of Income Division providing the remaining records from a project involving estate tax returns.

Match to the National Death Index

The next step in the process was the use of the National Death Index (NDI), to obtain the date of death and the death certificate number. To prepare for the match, we updated the SOI file by obtaining the following information from the IRS's Individual Master File: (1) name, (2) most recent State of residence, and (3) most recent marital status. We also passed the file against the SSA's Year of Birth File to obtain the month, day, and year of birth for each taxpayer in the 1979 SOI file. This means that we have the age of the taxpayer whether or not he or she matched to the NDI. This information has been used to compute death rates by age, and to compare them to expected death rates.

The updated file was matched to the NDI. A total of 354,254 records were used for this match (one per taxpayer--duplicate returns for the same taxpayer were eliminated from the match). Out of the 354,254 records, 121,708 records were found to be possible matches to the NDI. At first glance, it appeared that one third of our taxpayers had died within a 5-year period, and many of these had died three or four times. A closer look revealed that only 4,056 of these records matched exactly on all items (first name, last name, SSN, birth date, sex, marital status, and State of residence). The remaining 117,652 records matched to the NDI on some but not all of these items.

Once we received the output from the NDI match, each record was assigned a score. Points or weights were assigned to each item that matched with the NDI; a perfect or exact match received a score of 100. A score was devised to be assigned to each item which matched between the two files. The scores achieved in each of these comparisons were added to each other to obtain an overall score for the record. In addition to the score, we assigned a three-digit "hit-type" code to each record for ease of tabulation and analysis.

Tabulation of the data by score and hit-type led to some obvious "bunching" of death certificates--ranges of scores for which we had a large number of records, ranges where there were very few. The lower end of the top

"bunches" suggested themselves as logical cut-off points for ordering death certificates from the States. We used a specific algorithm which called for the ordering of all death certificates which had partial matches because of missing data from one or both files--such as SSN's which may have been off by one or two digits, no date of birth match, or no last name match, etc.

We compared the results of our selection process to the total number of taxpayers in our original sample and computed a "tentative death rate" ("tentative" because this was just a preliminary selection of records; a manual selection was yet to be made after the records were received from the States). A comparison was then made to "expected" death rates, based on the age distribution of our sample and life tables from the National Center for Health Statistics.

Since the number of records selected under this algorithm (21,078) exceeded the number of expected deaths for our sample (15,239), we felt fairly confident that we were ordering enough death certificates to get nearly all of the true matches.

Obtaining the Death Certificates

To obtain the death certificates, the SOI Division followed the instructions given in a publication prepared by the U. S. Department of Health and Human Services entitled "The National Death Index File Search: Interpreting the Output" [14]. Introductory letters were prepared in March 1985, and sent to each of the designated persons in each State informing them of our plans for the study. By late summer, we had heard from most of the States. Most of them were in agreement with our research.

In September, 1985, letters were sent to all States that had not turned down our project, along with a computer printout showing the State name, date of death, and death certificate number for each decedent selected for the study. The States were asked to pull the requested death certificates by matching on death certificate number. Follow-up letters were sent to each State that raised objections to our original request, as well as those that did not respond to our September request. Nearly two years elapsed between the time we sent our original requests to the States and the time we received our final shipment of death certificates.

Some of the States objected to the Foreign Branch's study, which involved computation of the estate multiplier (a formula by which the wealth of the living is estimated from the wealth of the deceased) [15]. The second problem, involving four States, took even longer to resolve. These States insisted on receiving the names and social security numbers of the alleged decedents so that they could verify whether the records they were giving us truly belonged to the individual in the study. Since our Disclosure Division would not agree to let us release this additional information, we worked out the following compromises: one of the four States agreed to pull the records without receiving names and SSN's if we eliminated all records that did not match to the NDI on

name and SSN. This meant that we could only send to that State a pull-list with scores far exceeding our cut-off point. The three remaining States allowed us to travel to their offices of vital statistics, pull the records ourselves, and perform a verification process that assured the States that we had pulled the proper record.

As the death certificates were received in the SOI Division, they were compared with computer listings which contained all the relevant information we had on each taxpayer--name, birth date, SSN, sex, marital status, spouse's name (if a joint return), and State of residence. As was mentioned earlier, during the regular 1979 SOI processing, our editors were asked to enter a sex code on the return based on the first name of the taxpayer. We were delighted to find only 32 cases where the "sex" code placed on the tax return, based on the taxpayer's name, disagreed with the sex of the decedent shown on the death certificate.

Death certificates received from the first five States were reviewed and analyzed by the SOI Division's professional staff. On occasion, we found that the wife had used her deceased husband's SSN, especially among the elderly. A detailed analysis of the records from several States suggested a clerical operation by which the 21,078 death certificates could be reduced to 11,041 [16]. Nearly 1,000 more were deleted during a subsequent professional review. The observed death rates were then compared to the expected death rates (see Table 2).

Obviously, death rates for taxpayers in our sample do not increase as rapidly with age as do death rates for the population as a whole. Three factors may be at work here. First, taxpayers as a group are presumably healthier than the population as a whole, and therefore have lower death rates within each age group. Second, we may have under-selected death certificates of aged taxpayers because of problems with their use of SSNs [17]. However, the major difference between the population and taxpayer death rates lies in the 65 and over age group. It would appear that a significant proportion of the elderly escaped our mortality study by dropping out of the tax-filing population more than five years before their deaths. It should be noted that the shortfall in elderly, retired individuals should not have a major effect on occupation-based mortality data.

Once we had selected death certificates for further processing--those determined to be "true matches"--we masked all identifying information, photocopied them, and sent one photocopy to the National Center for Health Statistics (NCHS) for coding the cause of death, and one photocopy to the Census Bureau for coding the industry and occupation of the deceased--the occupation as shown on the death certificate.

Coding the Death Certificates

The death certificate has a space for the physician or medical examiner to record the cause of death. It also has a space for the underlying cause of death which is to say the condition that started the sequence of events which went from a normal healthy person to the immediate cause of death. For example, a decedent's underlying cause of death may have been

lung cancer but the immediate cause of death may have been pneumonia or heart failure. This decedent would be counted with decedents dying of lung cancer. The National Center for Health Statistics coded the death certificates by using a 4-digit code for each of the causes of death classified according to the International Classification of Diseases [18]. The total number of U. S. deaths between 1979 and 1983 classified by underlying cause of death are shown in Table 3.

LINKING THE OCCUPATION-CODED FILE TO THE MORTALITY FILE

The data from the Occupation-Coded File were linked to the Mortality File by matching the files on a unique control number. Mortality rates by age, marital status, sex, and cause of death will be tabulated for major occupations within industries for NCI. After all data have been tabulated and reviewed, a public-use tape of unidentifiable data will be prepared for NCI to allow them to relate death rates and causes of death to occupation and industry.

CONCLUSION AND FUTURE USES OF THE LINKED FILE

The IRS will compare the occupation as reported on the death certificate with that shown on the tax return. This comparison will enable us to help evaluate reporting on these forms. Also, we plan to prepare some data on achieved death rates by occupation and income class.

We could code a much larger sample, at a much lower cost, if the keying of the occupation could be integrated into our regular revenue processing system. We could pre-designate a sample by Social Security Number ending digits, and key the taxpayer entries every time the new input system (currently being developed for IRS) identifies one of these endings. We plan to use our Computerized Occupation Dictionary to code the entire Tax Year 1989 SOI individual sample. Using the computerized dictionary, the coding should go much faster the second time around.

We believe the 1979 Occupation Coded File has already proven to be an invaluable tool with the enhancements we have given it. We will now have to analyze the Mortality data, i.e. mortality rates by exposure groups, causes of death by occupation/industry, etc. Even if the sample proves too small to provide statistically significant differences in mortality rates by exposure groups, we will at least have proved that it is possible to get the necessary clearances to do this kind of study, and will have laid the groundwork for future large-scale studies.

ACKNOWLEDGMENTS

Many individuals have contributed to the research done over the ten-year life of this project. At the risk of leaving out some of them, the authors would like to thank Dr. Gilbert Beebe (NCI); Dr. Harriet (Orcutt) Duleep (then of SSA); Phil Clark and Patricia Crabbe (then of IRS); and Beth Kilss and Dr. Fritz Scheuren (IRS) for their many contributions.

Table 2.--Number of Deaths between 1979 and 1983
by Underlying Cause of Death

Underlying Cause of Death	Number from SOI	Number from NCHS ¹
Total Deaths.....	4,174,213	9,875,661
Infectious and Parasitic Diseases, Total	26,817	92,461
Tuberculosis.....	995	9,508
Bacterial Diseases.....	15,268	57,888
Viral Diseases.....	5,564	11,083
Others.....	4,990	13,982
Neoplasms, Total.....	1,060,975	2,150,054
Cancer of the Pancreas.....	49,508	108,711
Lung, Trachea, and Bronchus Cancer.....	290,571	535,272
Female Breast Cancer.....	80,806	181,917
Cancer of the Prostate.....	60,886	117,491
Other Cancers.....	579,204	1,206,663
Endocrine, Nutritional, Metabolic Diseases, & Immunity Disorders, Total... ..	63,328	230,884
Diabetes Mellitus.....	45,986	173,574
Nutritional Deficiencies.....	1,722	11,991
Obesity.....	7,472	5,089
Others.....	8,178	40,290
Mental Disorders, Total.....	19,604	71,696
Alcoholic Dependence Syndrome.....	7,772	20,578
Drug Dependence.....	4,431	3,219
Others, such as Senility, Drug and Alcohol Abuse, etc.	7,401	47,899
Diseases of the Circulatory System, Total.....	1,750,454	4,892,763
Hypertensive Diseases.....	60,699	159,450
Ischemic Heart Diseases.....	1,058,870	2,777,576
Cerebrovascular Diseases.....	223,873	816,525
Others.....	407,012	1,139,212
Diseases of the Respiratory System, Total.....	208,554	642,293
Pneumonia.....	67,185	249,650
Obstructive Pulmonary Conditions.....	104,869	290,930
Others.....	36,500	101,713
Injuries and Poisonings, Total.....	494,639	767,057
Motor Vehicle Traffic Accidents.....	110,440	243,762
Suicide.....	30,879	138,208
Homicide and Legal Intervention.....	53,603	113,023
Other Accidental Injuries and Poisonings.....	299,717	272,064
All Other Causes of Death Not Shown Above.....	549,842	1,028,453

¹ Includes infant, preadolescent, and adolescent deaths, as well as deaths of older Americans not included in the taxpaying population. Source: Vital Statistics of the United States, Volume II - Mortality, Part A, U.S. Department of Health and Human Services, Public Health Service, National Center for Health Statistics, Tables 1-23, 1979-1984.

Table 3.--Five-Year Death Rates by Age and Sex, 1979-1983

Age	U.S. Population ¹ 1979	U.S. Taxpayers 1979
Total.....	4.5	3.0
Under 25.....	.6	.6
25 - 34.....	.7	.7
35 - 44.....	1.2	1.1
45 - 54.....	2.8	2.8
55 - 64.....	6.9	6.1
65 and over.....	27.2	15.5
Age not specified.....	N/A	1.6
Total, Males.....	5.7	3.9
Under 25.....	.7	.9
25 - 34.....	1.0	1.1
35 - 44.....	1.5	1.4
45 - 54.....	3.6	3.7
55 - 64.....	9.2	8.0
65 and over.....	32.8	19.9
Age not specified.....	N/A	1.4
Total, Females.....	4.0	2.1
Under 25.....	.4	.4
25 - 34.....	.4	.4
35 - 44.....	.8	.7
45 - 54.....	2.0	1.9
55 - 64.....	4.9	4.3
65 and over.....	23.4	11.3
Age not specified.....	N/A	1.7

¹ Source: U. S. Bureau of the Census, U.S. Census of Population: 1960, 1970 and 1979, Volume 1; and 9th Revision of International Classification of Diseases. The rate was calculated by dividing the number of deaths by the population in 1979.

NOTES AND REFERENCES

- [1] Sailer, Peter, Orcutt, Harriet, and Clark, Phil, "Coming Soon: Taxpayer Data Classified by Occupation," 1980 American Statistical Association Proceedings, Section on Survey Research Methods, pp. 467-471.
- [2] U. S. Internal Revenue, Statistics of Income for 1916; 1918, p. 13 and pp. 98-137.
- [3] On a more serious note, 1979 was chosen by us in part because the Census results would be available for comparison. Comparisons could also be made between Current Population Survey (CPS) and IRS data as was done in 1963, but it was judged better to look at the Census, since, like the IRS data, it is self-reported.
- [4] It should be noted, however, that errors can occur in the industry information obtained from the social security employer file as described by David W. Cartwright in "Major Limitations of CWS Files and Prospects for Improvement," Policy Analysis with Social Security Research Files, Social Security Administration, Research Report No. 52, 1978, pp. 580-581. See also, Levine, Bruce, "Improving Industry and Place of Work Coding in the Continuous Work History Samples," 1980 American Statistical Association Proceedings, Section on Survey Research Methods.
- [5] U.S. Department of Commerce, Bureau of the Census, 1980 Decennial Census Processing Manual, Volume V, Chapter 5, "Industry and Occupation Coding," Section D, "Production Coder Procedures," June 20, 1980.
- [6] U.S. Department of Commerce, Bureau of the Census, 1980 Census of Population: Alphabetical Index of Industries and Occupations, Second Edition, May 1981.
- [7] U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, Standard Occupational Classification Manual, 1980.
- [8] U.S. Department of Labor, Employment and Training Administration, Dictionary of Occupational Titles, Fourth Edition, U.S. Government Printing Office, 1977.
- [9] Appel, Martin and Hellerman, Eli, "The Census Bureau Experience with Automated Industry and Occupation Coding," 1983 American Statistical Association Proceedings, Section on Survey Research Methods, pp. 32-40. For experiences with automated or semi-automated occupation coding in other countries, see Lery, Alain and Stephany, Arnaud, "COLIBRI II: Saisie-Chiffrement en Ligne des Bulletins du Recensement de la Population de 1982," Courier des Statistiques, No. 30, April 1984, pp. 20-28; see also Lyberg, Lars, "The Development of Procedures for Industry and Occupation Coding at Statistics Sweden," Statistical Review 1983, No. 5: Essays in Honour of Tore E. Dalenius, pp. 139-156.
- [10] U.S. Department of Commerce, Bureau of the Census, 1980 Census of Population: Detailed Occupation and Years of School Completed by Age, for the Civilian Labor Force by Sex, Race, and Spanish Origin: 1980, Supplementary Report, PC80-S1-8, March 1983.
- [11] U.S. Department of Treasury, Internal Revenue Service, Statistics of Income-1979, Individual Income Tax Returns, U.S. Government Printing Office, 1982. (For Tax Year 1979, filing requirements ranged from \$3,300 for a single person under age 65 to \$7,400 for a married couple age 65 or over.
- [12] The Census adjusted occupation distribution in Table 1 is based on tables produced for the 1980 Census Public-Use Microdata Sample by DUAL-COMM, Inc. under the direction of John Beresford.
- [13] To make a comparison between the IRS and Census data, the IRS and Census occupation codes had to be made equivalent. Thus, Census occupation codes were recoded to the major group of two-digit SOC level.
- [14] The National Death Index File Search: Interpreting the Output, Department of Health and Human Services, Public Health Service, National Center for Health Statistics, Revised August 1, 1985.
- [15] Several States felt this was not an appropriate use of death certificate information. For these States, we removed the Foreign Branch's records before sending the pull-lists to the State.
- [16] If the SSN did not match, we accepted all matches with scores greater than 74; If the SSN matched, but the year of birth and first name did not, we accepted all matches with scores greater than 89; All remaining records with SSN matches and with scores greater than 82 were accepted.
- [17] Problems on the death certificates, as well as in SSA's name file, which may have caused us to get the wrong name (and therefore a low match score) for some of these aged taxpayers.
- [18] International Classification of Diseases. Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death. Based on the recommendations of the Ninth Revision Conference, 1975, and Adopted by the World Health Assembly, Volume 1, Geneva: World Health Organization, 1977.