

The Social Security Administration's Continuous Work History Sample

by Creston M. Smith*

In 1939, the original Social Security Act was amended to require that eligibility for benefits and payment amounts would be determined based on quarters of coverage (employment) and annual wages. These amendments required the Social Security Administration (SSA) to develop a recordkeeping system that collected quarterly wage and salary amounts for workers covered by the Act. The Continuous Work History Sample (CWHS) evolved from the early production of information from this employment data base. Through the years, the CWHS has adapted to changes in Social Security programs and linkages were developed to employer and benefit data bases. The CWHS is a major source of Social Security program statistics and workforce data. Tabulations and microdata files from the CWHS data bases are regularly used by SSA staff in making revenue estimates, evaluating legislative proposals, and responding to informational inquiries.

Today, the CWHS data in aggregated form are used at various levels of government outside of SSA for special analytic and research needs. In the past, special versions of the CWHS microdata files were made available to researchers under strict confidentiality restraints and with signed agreements on the specific conditions of use. However, certain confidentiality and disclosure issues raised in the interpretation of the Tax Reform Act of 1976 have prevented any recent releases of the CWHS to non-SSA researchers.

*Division of Statistical Operations and Services, Office of Research and Statistics, Office of Policy, Social Security Administration. The author wishes to acknowledge the substantial contributions of Lois Alexander, Warren Buckler, Barry Bye, and John DiMaggio.

The Continuous Work History Sample (CWHS) is the oldest major longitudinal sample data source in the Federal statistical system. It was developed to fulfill the need for statistics to be used in planning and operating the old-age insurance system established by the Social Security Act of 1935.

The provisions of the Social Security Act made possible the development of comprehensive and accurate data on the continuity of earnings and employment. The Act required employers to file quarterly reports of individual workers' taxable wages that were subject to the Federal Insurance Contributions Act (FICA)—these data, in turn, provided the base for a system of wage recordkeeping. Total taxable wages received by an individual from various employers could be determined and patterns of earnings and employment of a worker over an extended time could be traced.

It was recognized at an early stage that wage records had value for broad economic and demographic research purposes as well as for administering the old-age insurance system, which was the responsibility of the Bureau of Old-Age and Survivors Insurance (BOASI), the predecessor of the Social Security Administration (SSA).

In 1938, BOASI began tabulating wage record data on 1937 annual earnings by age, sex, race, and the

State in which the worker was employed. These annual tabulations were repeated for 1938 and 1939. Additionally, the data for 1937-39 were tabulated by industry and cumulative wages. The 1937 and 1938 data were tabulated on a 100-percent basis; data for 1939 were based on a 20-percent sample. When, in 1939, the Social Security Act was amended to require continuity of employment and wages as a basis for determining benefit eligibility and amounts, the type of statistics necessary to carry out the provision differed from what had been collected earlier. The number and specified quarters of coverage were now needed along with the cumulative wages. The first full national sample was intended to include 4 percent of all living wage workers whose accounts were established during 1937-40.

Selection was made on the basis of certain specified digits of the Social Security Number (SSN). The current year's wage record for each worker in the sample was to be added to the accumulated information, and the accounts of deceased workers and those who became entitled to primary insurance benefits were to be segregated from the others. Because of the wartime need for economy in the use of personnel time, the sample was limited to one-fourth of the originally planned size (1 percent).

The **Social Security Bulletin** reported on the progress of and

findings from the CWHS, beginning with a technical description of the CWHS in February 1944. A 1951 article presented summary statistics and an analysis of the data from the 1937-48 CWHS, a 12-year accumulation of data under that system.¹ Through the years, the CWHS system has been adapted to keep pace with the numerous changes in the Social Security program. As the Social Security Act was amended, the CWHS processing, files, and record contents were modified to reflect the changes.² This article discusses the current structure of the CWHS data system and files and describes some uses of the sample data and a number of issues that will affect the future of the system.

Sample Design

The 1-percent sample may be described as a stratified cluster

¹ See Jacob Perlman and Benjamin Mandel, "The Continuous Work History Sample Under Old-Age and Survivors Insurance," **Social Security Bulletin**, February 1944, pages 12-22, and Jacob Perlman, "The Continuous Work History Sample: The First 12 Years," **Social Security Bulletin**, April 1951, pages 3-10.

² Warren Buckler and Creston M. Smith, "The Continuous Work History Sample: Description and Contents," in **Policy Analysis Social Security Research Files** (Research Report No. 52), Office of Research and Statistics, Office of Policy, Social Security Administration, 1979.

probability sample of all possible SSN's.³ The population, or sampling frame, from which the CWHS cases are selected consists of the 1 billion possible nine-digit SSN's. These digits represent:

Geographic area of assigned number (three digits)	Time of issue group number (two digits)	Random serial number (four digits)
XXX	XX	XXXX

In issuing SSN's, each State is assigned one or more area numbers (first three digits).⁴ The fourth and fifth digits form a group number that provides a rough measure of when a batch number was released for issuance. Together the area number and group number define a stratum of the population of possible numbers. The numbers are thus stratified geographically by place of application for SSN and chronologically by the process of assigning these numbers within each stratum.

The actual sample consists of all persons who have SSN's with specified digits in certain of the serial-number positions (digits 6-9) of the SSN. The sample is clustered because of the particular digital selection procedure used. The clustering has been found to have some effect on the sampling errors of estimates from the 1-percent CWHS. The errors for some estimates, although in an accepted range, are slightly larger than those that would result from a stratified random sample of the same size.⁵

³Leslie Kish, *Survey Sampling*, New York: John Wiley & Sons, 1965.

⁴A special block of numbers was assigned to workers covered under the Railroad Retirement Act.

⁵Robert H. Finch, Jr., *Sampling Variability in the 1-Percent Continuous Work History Sample*, Office of Research and Statistics, Social Security Administration, 1977.

Data Sources

Demographic and Earnings Data

Most data for the CWHS are obtained from records derived from the reporting and information forms used in administering SSA's retirement, survivors, and disability insurance programs. The date of birth, sex, and race of the individual are obtained from the Application for a Social Security Number (Form SS-5). The CWHS draws on the administrative records created by the earnings posting process to obtain wage data for individual workers. These data are reported by employers using the Transmittal of Income and Tax Statements (Form W-3) and Wage and Tax Statement (Form W-2). In a massive operation, SSA processes approximately 200 million individual Form W-2 wage items annually. These items are balanced to employer-supplied totals and validated to existing names and SSN's on file at SSA. The CWHS has the advantage of the accounting process checking the wage data before it is extracted for statistical purposes.

Before 1978, a wage quarter of coverage was accrued for each calendar quarter in which \$50 or more were earned by working at a job covered by Social Security, and wage information before 1978 was collected on a quarterly basis. Starting with the Annual Wage Reporting (AWR) in 1978, quarters of coverage were determined based on a set amount of FICA covered taxable earnings. In 1978, a quarter of coverage was earned for each \$250 of covered earnings, up to a maximum of 4 quarters. The quarter-of-coverage dollar equivalent is indexed each year; in 1989, the amount was \$500 per quarter of coverage. The old method of quarterly reporting had its own associated series of reporting forms.

Regular wage employees were reported on Form 941, household employees on Form 942, and State and local government employees on OAR-S3. Farm employees' wages were reported annually on Form 943, and self-employed persons used Schedule SE of Form 1040 to report annually. With the maturing of the Annual Wage Reporting process, the W-3 and W-2 reporting forms have replaced most of the special coverage forms. The one exception is the self-employed, who still report on Schedule SE of Form 1040. Earnings from self-employment are processed first by the Internal Revenue Service (IRS) and the FICA earnings information is then supplied to SSA on magnetic tape.

Employer Data

Geographic and industry information is obtained from the employer's Application for an Identification Number (Form SS-4) and other related forms that are used periodically to clarify and update this information. Initially, employers are assigned geographic and industry classifications based on the location and nature of business information supplied on the Form SS-4. The Form SS-4 has also been used to develop information on employment that over the years has been exempt from FICA coverage or has been subject to special coverage or reporting requirements. The major special coverage groups classified are agriculture, Federal civilian, State and local government, household, nonprofit, and military.

Early in the development of the CWHS, arrangements were made to obtain additional information on large diversified business entities. At the present time, no attempt is made to collect detailed business

activity or geographic information on the Form SS-4. Employers who operate more than one place of business and have a total of 100 employees or more with at least 6 of them in a separate location are asked to use the Establishment Reporting Plan (ERP). Under this plan, the employer gives SSA a list showing the location, industrial activity, and approximate number of employees of each reporting unit, and identifying each group with a preassigned establishment number. This arrangement allows SSA to more precisely classify employees by geographic location and industry. There are indications that employer participation in ERP has declined in recent years. This issue is discussed in more detail below.

Claims Data

The CWHS was created in response to the need to study the work, earnings, and employment data that were the tax base for funding the Social Security programs. It has continued to concentrate on the employment and coverage aspects of these programs. However, the convenient size of the sample led to the incorporation of certain key information from the claims and benefit side of the program. The addition of claims information has allowed program analysts to conduct special studies of beneficiaries and their work histories. The claims information is drawn from SSA's Master Beneficiary Record (MBR), which is a compilation of information collected at the time an individual files a claim with SSA and the claimant's subsequent benefit history. The beneficiary data in the CWHS gives dates of entitlement, benefit amounts, and type of benefit. The extraction of the data from the MBR is used to construct a point-in-time observation

of benefits that were in force for December 31 of the particular data year.

Special CWHS Data Fields

Two special data fields are created in the CWHS update process that are not part of the parent administrative record systems: Insurance status and benefit status. Insured status measures an individual's projected eligibility for retirement and entitlement to survivor benefits based on the number of quarters of coverage accrued thus far in his or her work history. The codes in this field indicate if sufficient quarters of coverage have been earned to guarantee benefit eligibility or for younger workers if quarters of coverage are being earned at a rate that will result in guaranteed benefit eligibility at the attainment of age 65. The benefit status field shows whether or not benefits are being received under the sample SSN. The benefit status code also identifies the type of benefits being paid—retirement, survivors, or disability benefits. Both codes are available for each calendar year, which permits analysis of change of status over time.

Statistical Processing and Data File Contents

The 1-percent CWHS files are processed on an annual cycle. The cycle begins with the extraction of the 1-percent earnings data from the Earnings Reference File (ERF). This extraction process begins in December of the current year to collect data for the previous tax year's wage items. The earnings extraction also updates all previous years' earnings for corrections or adjustments that have been processed by SSA during the year. The ERF extraction becomes the

foundation of the CWHS update. The summary portion of the ERF is used with the existing 1-percent 1937-to-date CWHS file and the 1-percent file extracted from the MBR to generate an annual update to the 1-percent 1937-to-date CWHS file.

In addition to the annual summaries of taxable earnings, the ERF contains detailed accounting of postings made to each SSN after 1978. The ERF extraction obtained for updating the CWHS has the detail records. Each W-2 posted to the ERF summary and each Self-Employed (SE) report received from the IRS has a detail record. The records that apply to the current year's CWHS update are selected from the extract and matched to the employer coding files to obtain industry, geography, and special coverage codes. This file is the basis for the employee-employer data series.

Files Produced in CWHS System

A brief description of the files produced in the CWHS system follows.

One-percent 1937-to-date CWHS file.—A 1-percent sample of SSN's issued through the cut-off date of the file reflecting the individual's entire work experience in covered employment.⁶ Basic data elements include: (1) demographic characteristics—year of birth, sex, and race; (2) employment—number

⁶ A 0.1-percent sample for the period 1937-77 contains more detailed earnings data than the 1-percent CWHS. The detailed earnings information includes annual taxable wages for 1937-77; annual taxable farm wages for 1955-77; quarterly wages for each quarter of 1951-77; and annual net earnings from self-employment for 1956-77. Due to the discontinuance of quarterly wage reporting, the preparation of this file annually was discontinued after 1977.

and pattern of years employed, first and last years employed, and number of quarters of coverage from 1937 to date; (3) type of work—farm or nonfarm, wage, or self-employment; (4) taxable earnings each year from 1951 to date; (5) self-employment taxable income each year from 1951 to year prior to current year; (6) self-employment net earnings, for year prior to current year; (7) insurance status; (8) benefit status and dates of entitlement; (9) Medicare qualified government earnings (1983 and later); and (10) total compensation from the W-2 form (1978 and later).

One-percent sample annual Employee-Employer (EE-ER) file.—A 1-percent sample of SSN's for which wage and salary employment was reported in the reference year. One record exists for each employee-employer combination. Basic data elements include: (1) personal characteristics—year of birth, sex, and race; (2) wages—annual taxable, quarterly taxable (prior to 1978), total estimated wages (prior to 1978), and total compensation (1978 and later); (3) employer—State and county, industry, and coverage group (such as farm, household, and Federal civilian); (4) insurance status; (5) benefit status; and (6) Medicare qualified government earnings (1983 and later).

Each year, additional data files are also prepared that summarize the annual employee-employer data by major job and major geographic and industry categories. The basic definitions used in creating these auxiliary files are:

Assigned. Contains one record for each wage earner who earned at least \$1 in covered employment during the year. Geographic, industry, and coverage codes are "assigned" based on employer

of major job (employer with greatest taxable wage or total compensation).

Actual State. Contains one record for each State in which the worker had employment during the year. County is coded for employer of major job within State.

Actual Industry. Contains one record for each industry (2 digits) in which the worker had employment during the year. Third and fourth digits of industry are coded for employer of major job in the same industry code.

One-percent sample 1957-77 Longitudinal Employee-Employer Data (LEED) file.—A file assembled from the 1-percent sample annual EE-ER records prepared annually. In the annual files, one record is created for each employee-employer combination during the year. In the longitudinal file, the original records from the various annual files have been skeletonized, resequenced, and merged so that all records associated with an employee over the time span of the file appear together. Basic data elements are the same as in the 1-percent sample EE-ER records for the corresponding years. Planning for an update to the LEED for years after 1978 is described below.

One-percent sample annual Self-Employed (SE) file.—A 1-percent sample of SSN's for which self-employment earnings subject to Social Security coverage were reported in the reference year. Basic data elements include: (1) demographic characteristics—year of birth, sex, and race; (2) self-employment—taxable income, net earnings, State and county, and industry; (3) taxable earnings (including wages, if any); (4) type of work—farm or nonfarm self-

employment (and wage indication, if any); (5) insurance status; and (6) benefit status.

Uses of CWHS Data

The Social Security Administration's Office of Research and Statistics (ORS) and Office of the Actuary (OACT) use CWHS data in the form of tabulations and microdata files for Trust Fund revenue estimates, evaluating costs and benefits of proposed legislation, and dissemination of a variety of statistics about the program. The information is incorporated in reports or used directly by the Office of the Secretary of Health and Human Services, the Office of the Commissioner, Congressional offices, and the Board of Trustees for the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds. Before the Tax Reform Act of 1976 (discussed below), considerable use was made of this information system by researchers outside Social Security in both government and the private sector. Outside of SSA, use centered on studies of workers' geographic or industrial mobility, regional workforce characteristics, and longitudinal earnings patterns.

ORS Uses

The Office of Research and Statistics uses the CWHS to provide data that serve the need for baseline information on SSA programs. Much of this information is published in the **Annual Statistical Supplement to the Social Security Bulletin**.⁷ This Supplement contains a comprehensive overview of Social

⁷ **Annual Statistical Supplement to the Social Security Bulletin**, 1988 (for example, tables 4.B1-4.B10), pages 127-136, Office of Research and Statistics, Office of Policy, Social Security Administration.

Security programs, including tabulations from the CWHS that provide detailed information about OASDI workers and their FICA taxable earnings longitudinally, by age, sex, and type of earnings.

The CWHS is the sole source at SSA for information on the geographic distribution of workers and wages. Geographic information is published at the State level in the **Annual Statistical Supplement to the Social Security Bulletin** and county level data are published in a separate publication, **Earnings and Employment Data by State and County**.⁸ Geographic information from the CWHS is also provided in the **State Statistics** data series prepared by ORS and the Office of Governmental Affairs (OGA). These single-page-per-State data sheets are used by SSA officials in public presentations and for other informational activities.⁹ The **State Statistics** series provides a national overview of the Social Security program and a State-oriented summary of claims, benefits, and earnings information for a particular State. Other information from the CWHS is published in the **SSA Executive Handbook** and the House Ways and Means Committee Report on social programs.¹⁰

⁸ **Earnings and Employment Data for Wage and Salary Workers Covered Under Social Security, by State and County, 1985**, Office of Research and Statistics, Office of Policy, Social Security Administration, 1986.

⁹ **State Statistics**, Office of Research and Statistics and Office of Governmental Affairs, Social Security Administration, March 1988.

¹⁰ **Executive Handbook of Selected Data**, Office of Information Systems, Social Security Administration, May 1988, and U.S. Congress, House Committee on Ways and Means, **Background Material and Data on Programs Within the Jurisdiction of the House Committee on Ways and Means**, 101st Cong., 1st Sess., 1989, Committee Print WNCPC: 101-4, pages 3-126.

Data from the CWHS are also used in the revenue estimation process. Some of the estimates are used in an ORS econometric model for estimating payroll tax liability from the Bureau of the Census Current Population Survey. The model allows SSA to analyze the distribution of liability under current or proposed financing alternatives. Another series of estimates from the CWHS is used to provide 10-year projections of OASDI tax revenues for the budget, for the Office of Management and Budget (OMB), for the Trustees Reports and for general policy analysis. The CWHS data are also used to determine the amount of funds transferred from the Trust Funds needed to offset the liability created for the General Fund by overpayments of FICA taxes. For this transfer process, the CWHS employee-employer level data are used to analyze what part of the overpayments are from multi-employer situations and what part are error-excess payments from a single employer. In the latter case, liability to the employer as well as the employee is involved.

The longitudinal aspect of the CWHS data permits special SSA program studies that require the analysis of change in the distribution of worker and firm characteristics over time. For example, the CWHS was the source for recent internal estimates of the effect on labor supply of the 1983 elimination of the retirement earnings test for workers aged 70 and 71.¹¹ The study looked at the presence and level of earnings before and after the earnings test was eliminated for those workers.

¹¹ **The Effects of Removing 70 and 71 Year-Olds From Coverage Under the Social Security Earnings Test**, unpublished estimates, Office of Research and Statistics, Office of Policy, Social Security Administration, August 1988.

Currently, the potential of the CWHS is being explored for research on the demand for older workers. One project would look at the age composition of workers in large, stable firms. By obtaining a representative sample of the workforce in these firms over a number of years, the demand for older workers in the past and as the labor force ages in the future may be better understood.

Actuarial Uses

The Social Security Administration's Office of the Actuary also makes considerable use of CWHS information. Annual tabulations from the CWHS provide a longitudinal data base of material for various actuarial projections. The OACT also requests several large disc storage files of the tabulated data that are used in a data-base management system. One use of this data base is the construction of insured worker projections. The projections are a mandatory part of the Trustees Report to Congress and are used by the Office of Management and Budget in the Presidential Budget submittal to Congress. In addition to the large set of tabulated data, OACT uses data from the CWHS files for ongoing functions, such as building projection models for benefit payments and responding to requests for estimates of cost and savings from proposed legislation.

Outside Research Uses

Prior to the Tax Reform Act of 1976, release of the CWHS micro-data files was a common way of providing outside researchers with unaggregated SSA earnings and employment data. The releases were done on a reimbursable cost-recovery basis and the released edition of the file was prepared with precautions taken to ensure that

individual identities could not be associated with information about individual workers or employers. The file contained an encrypted identifier to preserve confidentiality while maintaining the ability to link files from one year to the next. Additionally, the release of the files fostered independent research related to Social Security programs at no cost to the Government and provided feedback to the data managers at SSA evaluating the quality of the data.

The data were made available to government and private researchers under a user agreement that placed restrictions on the file. The data could be used exclusively for research and statistical analysis, and the agreement specified the particular purposes for which the data were to be used. Prior SSA approval was required for the release of the files or file extracts to other organizations. The agreement prohibited all efforts to match exactly records of individuals with data from other sources; but, it did not preclude statistical matches.

During this period, the outside use of the CWHS had grown to the point where annual user conferences were regularly held to allow researchers to exchange study results, computer technologies employed in the use of the files, and statistical methodologies used in their diverse research. Through these CWHS User Conferences, ORS attempted to encourage file uses that would foster research in areas of Social Security programs. Outside researchers had an opportunity to present their requests and to discuss quality control problems and CWHS enhancements.

The CWHS files were used regularly by a substantial number of agencies of the Federal Government, State and local

government agencies, colleges and universities, and private research groups. The Department of Commerce found CWHS data helpful, especially in intercensal years, for monitoring and evaluating the effects of economic events and policies on workers. Within that Department, the Bureau of the Census and the Bureau of Economic Analysis (BEA) used the file extensively to estimate migration and economic and demographic characteristics of the population and to analyze changes in the geographic distributions of income and the impact of geographic mobility on wages and employment at the regional level.

A number of other Government agencies also used the CWHS data. Among them were the Department of Labor, the National Cancer Institute, the Internal Revenue Service, and the Department of Agriculture.

Various offices in the Department of Labor found CWHS data relevant to their program needs. The Bureau of Labor Statistics was one of the largest CWHS users. It combined CWHS data with its own files to prepare data for annual publications on the earnings of workers categorized by industry, age, sex, and race and to analyze labor turnover and the impact of industry employment changes on job attachment and earnings growth. The National Institute for Occupational Safety and Health used CWHS data for epidemiological and other health-related studies.

Other users were State and local government agencies, for studies of such things as employment security planning, economic development, and unemployment insurance policy; universities and colleges, whose major interest was in labor-market research; and a substantial number of private research organizations,

who used the file to study such things as income, geographic, and industry mobility, for analysis of migration patterns, and conducting research on various aspects of social and economic measurement.

Current Statistical Policy Issues

The evolution of the CWHS files has been guided by the structure of SSA administrative records and the information needs and research policy at SSA. This section describes two factors that will influence the future role and structure of the CWHS as a data base. From an SSA perspective, the major factor that continues to affect the CWHS files is the change to annual wage reporting.

Annual Wage Reporting

The change to annual wage reporting (AWR) has affected the type of information collected and, therefore, the kind of program analysis that is possible. To implement any system change that has an impact on 5-6 million employers reporting wages for more than 100 million workers involves significant difficulty, and the 1978 annual wage reporting changeover was no exception. In the early years of the changeover, difficulties involved late posting, duplicate reports, and other processing problems. The parent Earnings Reference File was, of course, corrected as reporting problems were discovered, and these corrections are reflected in the 1937-to-date CWHS file because the latest version of the Earnings Reference File summary portion is incorporated into the annual CWHS updates. However, the annual Employee-Employer and Self-Employed files have no retroactive

correction and remain a point-in-time observation of the original posting. The LEED file updates also reflect the uncorrected postings.

The annual report has created some need to reconfigure the structure of the LEED file. Before 1978, quarterly wage amounts were carried in the annual Employee-Employer files, and these amounts were included in updates to the LEED. From 1978 on, only annual data are available and some changes in the structure of the LEED file record contents will be required to produce a file of all available years that is compatible with the current system's software. The basic plan is to archive the LEED data for 1957-77 with quarterly data intact and to create a new LEED containing only annual data from 1957 to date.

Another area affected by the annual reports is the collection of data for special coverage groups. As discussed earlier, these groups often had special reporting requirements and unique forms (such as IRS Form 943 for agricultural employees and IRS Form 942 for household employees) that were used in the wage reporting process. Under the annual system, these special-coverage groups are reporting under the W-3, W-2 process. The W-3, W-2 process identifies some special coverage by check-off boxes on the W-3. However, it appears that the check-off system has not been totally successful. In particular, the number of agricultural workers identified in the CWHS has dropped sharply since the conversion.¹²

Employers using the Establishment Reporting Plan also have been affected by the change to annual reporting. During the conversion, many employers dropped the establishment codes from their annual wage reports. Although some employers returned to establishment reporting after several years, the net effect since the conversion has been a decline in the number of employers using the plan. This decline may have a negative impact on the reliability of geographic and industry data that are regularly produced. Industry data may not be so severely affected by large employers who do not report by establishment. Some indications exist that establishments of a given firm are often involved in closely related industrial activity.¹³ Although the effect on the geographic distributions may be more significant, the solution to the geographic problem might be provided through the use of address of residence from the worker's W-2 wage reports. This approach is currently being explored.

Although the change to annual reporting resulted in several serious problems, the new system provided some additional information to SSA researchers. All W-2's (even for non-FICA employment) are processed by SSA, so that the CWHS has gained access to wage reports not previously available under quarterly reporting. The data on non-FICA employment has been especially important in estimating revenue flows that would result from statutory changes in coverage.

¹³ Warren Buckler, "Employer Reporting Unit Match Study (ERUMS): A Progress Report," in *American Statistical Association 1985 Proceedings of the Section on Survey Research Methods*, Las Vegas, August 1985, pages 434-437.

These noncovered wage reports led to some restructuring of the annual employee-employer data files, which lead to the creation of a second set of files that contains both covered and noncovered jobs.

The availability of total compensation in the annual reporting process makes accurate data on wages exceeding the taxable maximum available for the first time. Under the quarterly reporting system, Social Security taxes were collected only up to a taxable limit. If a worker reached the limit in a particular quarter from one employer, that employer did not need to report additional earnings for that quarter or for any subsequent quarter. Special procedures were used to estimate total wages based on the quarterly patterns. Now, the W-2 contains the total compensation field used by employers to report total taxable income paid to a worker. This total compensation field is collected as part of SSA's earnings posting process. The information is used as the base for the wage-indexing process that sets taxable maximum wages—the amount needed to acquire a quarter of coverage and the indexing applied to previous years' earnings in the benefit calculations. The total compensation is now included in CWHS files in place of the estimated annual wages.

Release of Data to Outside Researchers

As discussed above, during the 1960's and the early 1970's, it was possible for SSA to make the CWHS microdata files available to researchers in government and the private sector, with restrictions on use and disclosure. The data were released on an individual worker level with encrypted identifiers

¹² There is less reason to study the old special coverage areas because these individuals now earn quarters of coverage on the same basis as regular wage and salary workers; however, from a research view they continue to be of interest in terms of the levels of quarters of coverage and benefits they receive.

under a restricted-use agreement. These files supported research in a wide range of socioeconomic areas. The Tax Reform Act of 1976 changed the statutory rules for disclosure of data in the CWHS. Previously, SSA's disclosure of data, including information obtained from the Internal Revenue Service, was controlled by the Social Security Act and the disclosure regulations issued pursuant to it. The 1976 amendments placed new confidentiality requirements in the Internal Revenue Act. They restrict disclosure of information on tax returns, including W-2, W-3, and SS-4 forms. Since 1976, the Internal Revenue Code and Treasury regulations have preempted disclosure authority over tax return information obtained by SSA and have effectively precluded the release of CWHS data. Although the Internal Revenue Code provides for release of statistical data, the IRS has interpreted those provisions narrowly to allow release of tables and summaries, but rarely of microdata.

The loss of access to this file by the research community has not gone unnoticed. The National Research Council (NRC), for example, in a 1988 publication, **The Aging Population in the Twenty-First Century**, reviewed and evaluated various data sources that could be used to develop health care policy for the elderly in the next decade.¹⁴ In that report, the CWHS is noted as an "unmatched" source of longitudinal data on the elderly. The potential of the CWHS, if enhanced with mortality and health care benefit data, is also

discussed. The report recommends that SSA and IRS review the possibility of releasing CWHS data sets under user agreements.

During the past decade, several attempts were made by previous users, notably the Bureau of Economic Analysis (BEA), to obtain CWHS data under a restricted-use agreement and with less specific earnings detail, but the IRS has never concurred on the release of a file. Most recently, SSA, IRS, and the National Cancer Institute have been attempting to develop a statistical file that incorporates information from the CWHS into an existing IRS-NCI project. The project will combine cause-of-death information with occupation, industry, earnings levels, and some demographic information on a small sample of workers. The file can then be used at IRS, SSA, and NCI. Even if this approach succeeds in providing a data file for use by these participants, it may not provide a framework for the general release of existing files to outside researchers.

Support Components

Because of its complexity and size, the CWHS requires substantial support to accomplish the collecting, editing, and updating of the information available. Within SSA, the Office of Research and Statistics (ORS) manages the CWHS and develops technical specifications for editing and data definitions within the current system.

However, the CWHS would not exist without the cooperation of and input from a large number of offices in SSA. The Office of Central Records Operations processes the Form SS-4 for SSA needs and, during the processing, performs the classification of employers to provide the industry, geography, and coverage information available in the CWHS. The Office of Systems Design and Development maintains the earnings systems (Earnings Reference File) that is the basic starting point for all major updates to the file. The Office of Systems Requirements (OSR) prepares specifications for and oversees extraction of the sample records from the master SSA files. The OSR also writes procedures and oversees the collection of employer information in the employer classification files. The Office of Material and Information Resources provides the computer programming and systems support for the update maintenance of the CWHS data bases. In addition to the work each of these organizations performs in the compiling of the CWHS data, each of them provides valuable insight into the understanding and analysis of data created from the system.

¹⁴ **The Aging Population in the Twenty-First Century**, Dorothy M. Guilford (editor), Washington, DC, National Academy Press, 1988.