Development of the Continuous Work-History Sample in Old-Age and Survivors Insurance

by B. J. MANDEL*

A MUCH better understanding of economic activity and demography than is now possible could probably be obtained if statistics on individual work patterns were available for a representative segment of the population throughout their working life. Such data would also give more insight into the relationship of different work patterns to both work span and life span.

Collecting comprehensive data on work histories would require a study of work patterns for the entire working life of individuals. A survey extending over 30 or 40 years or more obviously requires long-term planning, fairly solid financial backing, and a well-established administrative and technical organization that can deal with all the problems of collecting the data. In the past, therefore, no data-collecting agency in the United States has directly undertaken such a comprehensive study. The establishment, however, of the old-age and survivors insurance program,[1] which bases a person's eligibility for benefits upon his lifetime earnings and work experience, has made possible as a byproduct of the wage records the beginning of a work-history statistical program that has vast potentialities.

The Bureau of Old-Age and Survivors Insurance has established and maintains on a continuing basis earnings records for more than 100 million individuals with wage or earnings credits under the program. When the wage records were being established in the early years of old-age and survivors insurance, it was recognized that they could become a valuable source of statistical information on many aspects of the social security program, such as coverage and the extent of protection afforded American workers and their families. It was also recognized that social and economic data on many other aspects of employment and earnings patterns could be derived from these records. As early as 1938, when data were tabulated on the first year's experience under the program in 1937, these expectations started being realized.

In 1938, for the first time, comprehensive data became available on the number of persons who were in covered employment at any time during the year. When this number was compared with the number of available jobs, an unexpectedly large difference between during-the-year workers and jobs was found. The ideas of many analysts about the extent of job turnover in a year were revised on the basis of these findings. After 1938 new concepts of employment were developed, more data were added to the records for the individuals originally covered under the program, and records were established for the many new entrants. For the years after 1937 statistical tabulations were made showing the number of workers employed regularly in a year and for a number of years, and data were tabulated on the earnings of different types of workers by such characteristics as age and sex. The statistics were used in evaluating proposals for legislative changes and in making various administrative plans and decisions.[2]

The Bureau of Old-Age and Survivors Insurance maintains a permanent 1-percent sample of workers, called the continuous work-history sample, which provides a source for much of the statistical information that it needs to carry on its research and administrative functions. The article that follows describes the sample; the types of data it provides; selection and maintenance of the sample; the potential role of electronics in processing the sample; and some practical problems in developing its statistical potentialities.

Composition of the Sample

The continuous work-history sample consists of a scientifically selected 1-percent sample of all persons who have a social security account number. The account number consists of nine digits and is issued to a worker on application; this number is the basis for identifying him throughout his working life and during the period when benefit payments are made to him or his dependents. Every person with specific digits in his account number is included in the sample. Once a worker is selected for the sample he remains in it permanently. Consequently, while as a percentage of all account-number holders the sample remains constant, the number of persons represented by the sample continues to increase. It currently includes more than 1 million account numbers.

For each worker in the sample, records of selected items of information about his work and earnings record since 1936 are maintained—his year of entry into covered employment, years of employment, amount of covered earnings, and benefit status (whether living and entitled to benefits or deceased). In addition, there is information on the worker's year of birth, sex, place of employment, and so on. Following is a list of the basic items that were included in the 1937–54 continuous work-history sample:

* Division of Program Analysis, Bureau of Old-Age and Survivors Insurance. The article is based on a paper prepared for the meeting of the First International Conference of Social Security Actuaries and Statisticians, held in Brussels November 3–10, 1956.

[1] See Victor Christgau, "Old-Age and Survivors Insurance After Twenty Years," *Social Security Bulletin*, August 1955.

[2] For an analysis of the sample for 1937–48, see Jacob Perlman, "The Continuous Work-History Sample: The First 12 Years," *Social Security Bulletin*, April 1951.

Account number
Sex
Month and year of birth
Race
Number of quarters of coverage, 1937–54
Number of quarters of coverage, 1937–53
Number of quarters of coverage, 1951–54
Pattern of quarters of coverage, 1954
Number of quarters of coverage, 1953
Number of quarters of coverage, 1952
Insured status, January 1, 1955
Insured status, January 1, 1954
Pattern of years employed
Number of quarters employed, 1954
Number of quarters employed, 1953
Years with $200 or more, 1937–50
Benefit status, January 1, 1955
Year of entitlement to benefits or of death
Quarterly wages, 1954
Taxable earnings, 1954
Taxable earnings, 1953
Taxable earnings, 1952 (in tens of dollars)
Taxable earnings, 1951 (in tens of dollars)
Cumulative earnings, 1937–54 (in tens of dollars)
Cumulative earnings, 1951–54
Coverage indication, 1954 (self-employed only, employee only, self-employed and employee both)
Coverage indication, 1953 (self-employed only, employee only, self-employed and employee both)
Card number

At present the continuous work-history sample provides data on work, covered earnings, and benefit status for the period 1937–54, and data are being developed for 1937–55. The term "work history" is not entirely an appropriate description of the present sample, since in covering only the first 18 years of operation of the program it includes only part of the working life span of most workers. It is planned to record the basic information for each worker in the sample for his entire working life. In the future, therefore, with the inclusion of work records covering 40 or more years, the sample will truly become a source of information on completed work histories.

Types of Data Provided

The continuous work-history sample was constructed to serve the needs of the Bureau of Old-Age and Survivors Insurance for data on program operations, and therefore the types of data tabulated from it largely reflect those needs.

One type of data derived from the sample—known as annual data—shows the number and percentage distribution of workers who received earnings credits for all 4 calendar quarters in selected years, classified by the amount of annual covered earnings. One of the many uses of these data is to show the average level of taxable earnings each year and provide a basis for forecasting future earnings levels. Another type of data—work-history data—shows how many workers have insurance status as a result of their accumulated work and earnings experience in covered employment and the patterns of years of employment for individual workers. Tables 1, 2, and 3 illustrate the types of data that may be compiled from the sample.

Both annual and work-history data are important in the operation of the program, since they show how many persons pay contributions and the amount they have paid, the number who were uninsured and their employment records, and the number of persons aged 60 and over and their work experience under the program. These data also provide a basis for estimating the Bureau's future workloads.

Selection of the Sample

The universe with which the work-history statistical program is concerned is made up of all individuals who have an account number. The aim is to get information about the covered employment and benefit status of all persons with account numbers under the program. The sample is sorted into two main groups: those who show no earnings credits, and those who had earnings credits at any time under the program. For the latter group, data are tabulated on accounts with earnings credits in the latest year and in previous years.

The system of sampling is based on the last four digits in the social security account number. The entire number consists of nine digits in three segments. The first segment (three digits) designates the geographic area where the account number is issued; the second (two digits) identifies the specific group or sequence of numbers issued in any one area; and the final segment (four digits) is the serial number.[3]

From within each area a 1-percent sample of the accounts is selected on the basis of specified digits in the serial number. A person who receives a number ending with 2505, for example, automatically is included in the sample. Altogether 100 predetermined digits out of the 10,000 possible in any one area and group fall into the sample. These 100 numbers all have a serial beginning with either a "2" or a "7." In other words, the sample is restricted to the blocks of

[3] For detail on the sampling methods, see the *Bulletin*, June 1951, pages 17–19.

Table 1.—*Number and percentage distribution of 4-quarter workers, by amount of annual earnings* [1] *in covered employment, selected years, 1939–54*

Amount	1939	1944	1946	1948	1950	1952 [2]	1954 [2]
	Number of workers (in thousands)						
Total	21,376	28,344	28,249	32,327	31,822	38,200	38,300
$1–999	9,327	4,519	3,531	2,752	2,261	2,650	2,600
1,000–1,999	8,877	9,288	9,921	8,017	6,899	6,700	5,950
2,000–2,999	2,149	7,442	8,408	9,572	8,836	8,850	7,950
3,000–3,599	425	3,233	2,890	5,319	5,316	5,000	4,750
3,600 or more	598	3,862	3,499	6,667	8,510	15,000	17,050
3,600–4,999	319	2,708	2,244	4,475	5,648	9,650	10,150
5,000 or more	279	1,154	1,255	2,192	2,862	5,350	6,900
	Percentage distribution						
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$1–999	43.6	15.9	12.5	8.5	7.1	6.9	6.8
1,000–1,999	41.5	32.8	35.1	24.8	21.7	17.5	15.5
2,000–2,999	10.1	26.3	29.8	29.6	27.8	23.2	20.8
3,000–3,599	2.0	11.4	10.2	16.5	16.7	13.1	12.4
3,600 or more	2.8	13.6	12.4	20.6	26.7	39.3	44.5
3,600–4,999	1.5	9.6	7.9	13.8	17.7	25.3	26.5
5,000 or more	1.3	4.1	4.4	6.8	9.0	14.0	18.0

[1] Includes wages in excess of the taxable limit. [2] Preliminary.

numbers that start either with 2000 or 7000 in the serial. This restriction was designed to yield a sample of accounts to which earnings for a full calendar year are posted at one time and at the earliest possible date. The accounts not in the sample have their earnings posted at later points of time for 4 calendar quarters that span 2 calendar years. In order that the sample might reflect complete calendar-year data, to fit in with most economic and statistical series, the accounts that had wage and earnings postings for a full calendar year were selected. For this reason the sample is stratified by area, and accounts in each area are selected systematically from selected clusters (the 2000 and 7000 series). The continuous work-history sample, therefore, is a two-stage sample.[4]

Each person who receives an account number has an equal chance—1 in 100—of falling into the sample. Thus, the sample has the important element of randomness as well as stratification and is kept current by the automatic addition of new accounts as they are issued. In 1951, for example, 4,927,120 new accounts were issued, and 49,495, or almost exactly 1 percent, had digits that fell in the sample. Each year approximately 1 percent of all newly issued accounts is added to the work-history sample. Because of sampling variation, precisely 1 percent is, of course, not attained.

Maintenance of the Sample

Each worker in the sample has two types of characteristics—fixed and varying. The fixed characteristics include such items as the worker's account number, date of birth, sex, race, and place of issuance of the account number. Examples of changing characteristics are earnings, work patterns, and benefit status.

Accounts of retired workers and deceased workers are identified but remain in the sample. Therefore, as new workers are included, there is a continuing accumulation of persons into the sample. The entire sample can be separated into persons entitled to benefits, those identified as deceased, and all others. It is possible to compile statistics on the employment history in any year for any of these three groups. At the end of 1955, the sample included more than 1 million accounts, representing 110 million persons who have had wage or earnings credits at any time during 1937–55. Of the 110 million, 4.5 million are receiving benefits based on their earnings accounts and 10.2 million are deceased. At some future date, perhaps when the program has been in operation for a period long enough to span an entire working life, it will be possible to tabulate life-span data for the accounts of deceased persons.

Maintenance of the sample also, by adding in the current year's work and earnings experience, brings up to date individual employment and earnings histories, so that a person's employment and earnings pattern are reflected in the latest record at the end of each year. From the current information, the varying characteristics of persons in the sample are determined—who and how many persons are new workers each year, who and how many have reentered or withdrawn from covered employment, and who and how many have been continuously employed over a period of consecutive years. The task of maintaining the sample is geared to a flow of basic records from the regular processes of keeping earnings and benefit records.[5]

Methods of Processing

The bulk of the processing operations is mechanized. Until 1956, punchcard equipment was used for bringing the individual records up to date and tabulating required statistics. The record of every person in the sample was placed on an IBM punchcard, and a numerical coding system was used to store information and permit economy in the use of punchcard space.

At first the punchcard method was satisfactory, but after several years the large accumulation of records for each individual in the sample and the limited space on the card, which has room for 80 single-digit codes, began to create problems. The number of different patterns of years of employment doubled each year. Though only one column was required to accommodate the possible combinations of years of employment in the first 3 years, two columns were needed when the sample covered 4–6 years; three columns when 7–9 years; four columns when 10–12 years and so on. Now, in the program's nineteenth year, seven columns are needed to account for all possible patterns of years of employment during 1937–55.

Under these conditions, if enormous increases in the statistical budget were to be avoided, the amount of work-history information had to be restricted. The situation was paradoxical. As a person's work-history record became more complete and more information on variations in work histories became available, the available equipment was less able to accommodate the requirements for such data. Consequently the statistical tabulations in time became limited to the most basic informational items needed for program administration and analysis. Such work-history items as changes in the insurance status of an individual from year to year or in his annual earnings capacity were not tabulated. Instead, data were obtained on the person's status at a given point of time—for example, his insurance status at the end of a year—or on his total cumulative or annual earnings, but not on his earnings experience and work patterns that led to that status.

Conversion to Electronic Processing

In 1949, when the earliest electronic computers became available, the Bureau began research into the possible use of this type of equipment for processing the work-history sample and concluded that it would be suitable. Before the decision was reached, a series of test runs was made with a small segment (100,000) of the continuous work-history cards and the

[4] For more detail see B. J. Mandel, "Sampling the Federal Old-Age and Survivors Insurance Records," *Journal of American Statistical Association*, September 1953.

[5] See William H. Cummins, "Old-Age and Survivors Insurance Records: Derivation of Byproduct Data," *Social Security Bulletin*, July 1952.

applicability and advantages of this equipment were definitely established. At present the entire 1-percent sample is being placed on electronic tape —the first step in using the main electronic computer. Instructions in electronic coding language are being prepared and will be programmed into the machine in order to tabulate information from the 1937–55 work-history record. It is planned to include the following 34 basic informational items in the 1937–55 records; some of them will be reconsidered in the light of cost and time factors.

Name
Account number
Indication of multiple account number
Sex
Race
Date of birth, periodically adjusted
Year of death
Source of death notice
Cumulative quarters of coverage, 1937 to date
Quarters of coverage, each of last 10 years
Quarters of coverage, each year since 1950
Year of last recomputation
Earnings credits each year, 1951–55
Self-employment income each year, 1951–55
Earnings credits each year, 1937–55
Year of disability onset and year of termination
Years with $200 or more, 1937–50
Cumulative earnings credits, 1937–55
Actual State and county, 1955
Actual industry, 1955
Actual class of work, 1955
Actual coverage class, 1955
Size of employer
Number of earnings items
Number of employers, 1955
Year of issue of account number
Estimated earnings in excess of earnings base, 1955
Benefit-in-force indication for old-age beneficiaries and deceased wage earners
Benefit status for wage earners
Year of entitlement
Primary insurance amount at end of year for old-age beneficiaries and deceased wage earners on whose records monthly death benefits are in force
Combination of family benefits in force, end of year
Amounts of family benefits in force, end of year
Summary of pattern of years employed, 1937–55

The basic information for 1937–55 represents a large increase from that for 1937–54. It would probably require several conventional IBM cards to store this information, and the costs of processing would be prohibitively expensive.

Table 2.—*Workers with earnings credits, work history: Number of 1937–53 workers, by age and sex, and percentage distribution by insurance status on January 1, 1954*

[Data derived from 0.1-percent sample and therefore subject to sampling variation that may be large where figures shown are small; include workers who died during the period. Data for male workers include workers of unreported sex. Age represents age at birthday in 1953. Figures in italics based on less than 100 workers. Data corrected to May 23, 1955.]

Age and sex	Number, 0.1-percent sample	Percentage distribution by insurance status, Jan. 1, 1954 [1]							
		Total	Fully insured			Uninsured			Deceased, benefit awarded
			Total	Permanently	Not permanently	Total	New entrants, 1953	Workers with previous wage credits	
Male	59,931	100.0	73.6	35.5	38.1	20.2	2.6	17.6	6.2
Under 20	3,592	100.0	30.9	[2]	30.9	69.1	28.8	40.3	[2]
20–24	5,525	100.0	69.2	*.1*	69.1	30.3	3.1	27.2	*.5*
25–29	6,544	100.0	83.8	5.0	78.8	15.1	*1.1*	14.0	*1.0*
30–34	6,731	100.0	82.3	23.2	59.1	15.1	.7	14.4	2.6
35–39	6,520	100.0	84.0	42.7	41.3	13.6	*.5*	13.1	2.4
40–44	5,892	100.0	84.4	49.5	34.9	12.5	*.8*	11.7	3.1
45–49	5,158	100.0	83.7	50.4	33.2	12.0	*.9*	11.1	4.4
50–54	4,475	100.0	80.2	54.5	25.6	13.4	*.6*	12.8	6.4
55–59	4,113	100.0	76.2	56.9	19.3	14.6	*.7*	13.9	9.2
60–64	3,548	100.0	70.9	62.3	8.6	16.5	*.9*	15.6	12.6
65–69	3,026	100.0	67.3	67.3	[4]	16.1	*.4*	15.6	16.6
70–74	2,101	100.0	57.6	57.6	[4]	19.2	*.4*	18.8	23.1
75 and over	2,180	100.0	37.2	37.2	[4]	26.5	*.2*	26.3	36.2
Unreported	526	100.0	*12.5*	*2.9*	*9.7*	86.9	*.2*	86.7	*.6*
Female	41,770	100.0	66.4	16.8	49.5	32.4	3.9	28.5	1.2
Under 20	2,691	100.0	27.8	[2]	27.8	72.1	32.7	39.5	[3]
20–24	4,917	100.0	65.8	[2]	65.8	34.1	4.0	30.1	*.1*
25–29	6,026	100.0	72.1	4.5	67.7	27.6	*1.1*	26.5	*.3*
30–34	6,101	100.0	71.0	11.3	59.7	28.4	*1.0*	27.3	*.6*
35–39	5,201	100.0	71.3	15.4	55.9	27.9	*1.7*	26.2	*.8*
40–44	4,386	100.0	71.6	18.3	53.3	27.3	*2.2*	25.1	*1.1*
45–49	3,625	100.0	72.0	22.5	49.5	26.8	*2.0*	24.9	*1.2*
50–54	2,865	100.0	67.9	29.5	38.4	30.3	*2.2*	28.1	*1.8*
55–59	2,191	100.0	65.6	37.2	28.4	31.9	*2.4*	29.5	*2.5*
60–64	1,565	100.0	65.4	51.4	13.9	31.4	*1.9*	29.6	*3.2*
65–69	1,100	100.0	61.1	61.1	[4]	33.8	*1.5*	32.4	*5.1*
70–74	601	100.0	58.7	58.7	[4]	33.3	*.3*	32.9	*8.0*
75 and over	382	100.0	41.1	41.1	[4]	42.9	*1.0*	41.9	*16.0*
Unreported	119	100.0	*12.6*	*5.0*	*7.6*	87.4	*.8*	86.6	[2]

[1] Except for workers on whose earnings records benefits were awarded and for whom the insurance status is the one determined at time of award, insurance status shown does not reflect changes in status arising from (1) combined earnings under the coordinated benefit provisions of the old-age and survivors insurance and railroad retirement programs, and (2) wage credits for military service.
[2] No workers in sample cell.
[3] Less than 0.05 percent.
[4] Not applicable under the Social Security Act.

Several major changes are expected from use of the electronic equipment. First, about ten times the information for one person can be stored with this equipment than with the 80-column punchcard. Another feature of the electronic equipment is its high rate of speed. It is estimated that some tabulations can be made about 100 times faster than with present punchcard equipment. The sorting speed is not so fast as the tabulating speed, but it is estimated in some projects to be 10–25 times faster than punchcard equipment.

In addition, the machine is expected to do a large part of the clerical work. The work of investigating discrepancies, preparing small summary tables from detailed tables, and calculating averages, percentages, and many other statistical measures will be transferred to the machine, which will also take over some of the final typing and graphic work.

Some Practical Problems

The task of selecting, maintaining, and processing the continuous work-history sample is only one aspect of the project. Equally important is the need to improve the quality and completeness of the basic data originating from the sample. A few of the problems are described briefly.

Identifying accounts of retired, disabled, and deceased workers.—Workers who have become entitled to old-age and survivors insurance benefits because they have reached age 65 are

readily identified through a cross check with a sample of retired beneficiaries whose account-number digits are in the 1-percent continuous work-history sample. Every year a match is made of the beneficiary and worker samples, and workers who have become beneficiaries are identified and their work history is coded.

In the course of this matching operation, workers who have filed a statement of total disability are identified. The matching process also identifies the persons who have died and on whose accounts lump-sum payments or survivor benefits have been awarded. It is this latter identification that provides the code on termination of the work history. Though this source completely identifies deaths among persons who were insured at the time of death and on whose account a claim was filed, deaths among uninsured workers are identified only in part. It is estimated that about a third of all persons with wage credits who died in 1955 were uninsured; approximately 80 percent of these deaths (about 25 percent of all deaths) were not reported to the Bureau. Thus a large gap in identification of deaths still remains. Matching against death certificates filed in the State health departments has been tried, but studies indicate that this is an expensive method. The main hope for identification is for the future when, because of the nearly universal coverage of the system, almost every worker with a substantial wage record who dies will be insured.

Patterns of years employed.—As previously mentioned, a basic classification in work-history studies deals with the specific combinations of years in which the person was employed. With this classification it is possible to study continuity and discontinuity in work patterns throughout the worker's life. When the tabulations for the years 1937 and 1938 were completed it was found that 15 percent of the persons worked in 1937 only, 11 percent worked in 1938 only, and 74 percent worked in both years. When the 1939 records were added, the number of combinations of employment patterns increased from three to seven; by 1955 the number had grown to 524,287. Although it was considered desirable to retain the classification "patterns of years of employment" in the sample, the lack of space on the punchcard made it necessary to condense the pattern to provide information on only: (1) the first year of covered employment, (2) the last year of employment, (3) whether employed in consecutive years, and (4) the total number of years of covered employment.

With the advent of electronic equipment it again became possible to include in the record all patterns of years of employment from 1937 on, and a reevaluation was made of the code. It was decided to maintain the old code of limited patterns for the period 1937–50 but that, beginning with 1951, when large numbers of additional persons were brought under the program, an effort would be made to maintain the complete pattern. The incompleteness of coverage in 1937–50 makes it impossible to determine without a further survey whether broken patterns are actually a result of a break in employment or of a shift to noncovered employment. With the further extension of coverage in 1957, under the 1956 amendments to the Social Security Act, nearly the entire labor force is covered, and the future patterns of years of employment will become more meaningful from the viewpoint of labor-market studies.

Multiple account numbers. — A problem that was serious at one time is that of "multiple" account numbers. The sample is a sample of account numbers, while the universe is one of individual workers. If a worker has more than one account number in the sample, he may be counted more than once. The problem was attacked in the early years of the program, when special efforts were made through a screening procedure to prevent issuing more than one number to an individual. In addition, informational campaigns were conducted to make covered workers realize that their records and earnings credits should be consolidated into one account to give them full protection and benefits. Multiple accounts are now consolidated at the time a claim for benefits is filed. From 1937 through 1955, approximately 5.8 mil-

Table 3.—*Workers with earnings credits, work history: Number of 1937–52 workers, by patterns of employment and insurance status, January 1, 1953*

[0.1-percent sample, tabulated data]

Pattern of years employed	Total	Fully insured	Uninsured	Deceased, benefit awarded
Total	[1] 98,378	68,656	26,015	3,707
Workers employed in 1952	58,101	49,748	8,187	166
Continuous patterns:				
16-year	9,888	9,818	1	69
15-year	451	446	1	4
14-year	579	576	[2]	3
13-year	701	700	[2]	1
12-year	1,095	1,090	[2]	5
11-year	1,372	1,367	1	4
10-year	1,300	1,296	1	3
9-year	896	894	2	[2]
8-year	789	785	3	1
7-year	989	982	6	1
6-year	941	930	10	1
5-year	1,023	984	38	1
4-year	1,056	961	95	[2]
3-year	1,827	1,387	439	1
2-year	4,675	2,461	2,209	5
1-year	3,484	[3]	3,484	[2]
Intermittent patterns:				
15-year	1,972	1,965	[2]	7
14-year	2,095	2,089	[2]	6
13-year	2,135	2,125	2	8
12-year	1,899	1,890	2	7
11-year	2,109	2,105	[2]	4
10-year	2,523	2,514	7	2
9-year	2,697	2,683	7	7
8-year	2,483	2,459	21	3
7-year	2,038	1,989	42	7
6-year	1,818	1,724	89	5
5-year	1,688	1,498	186	4
4-year	1,668	1,257	406	5
3-year	1,370	738	630	2
2-year	540	35	505	[2]
Workers not employed in 1952	40,277	18,908	17,828	3,541
Continuous patterns:				
15-year	394	294	[2]	100
14-year	344	253	1	90
13-year	393	260	1	132
12-year	455	310	1	144
11-year	563	401	1	161
10-year	699	515	1	183
9-year	903	710	4	189
8-year	1,001	814	9	178
7-year	1,235	990	17	228
6-year	1,614	1,287	55	272
5-year	1,883	1,472	134	277
4-year	2,393	1,691	424	278
3-year	2,917	1,439	1,235	243
2-year	4,321	490	3,644	187
1-year	8,217	[3]	8,110	107
Intermittent patterns:				
14-year	128	117	[2]	11
13-year	235	210	[2]	25
12-year	335	283	1	51
11-year	430	367	[2]	63
10-year	580	519	1	60
9-year	827	733	11	83
8-year	1,000	897	22	81
7-year	1,316	1,159	68	89
6-year	1,457	1,202	160	95
5-year	1,616	1,150	380	86
4-year	1,760	913	774	73
3-year	1,826	391	1,392	43
2-year	1,435	41	1,382	12

[1] Total excludes 10 workers for whom information is incomplete with respect to the characteristics reported here.
[2] No workers in sample cell.
[3] Not applicable; a worker must have at least 6 quarters of coverage to be fully insured.

lion multiple accounts representing about 2.8 million persons were consolidated and cross-referenced on the basic account records. Nevertheless, the problem is still present. A special study made in 1951 indicated that the continuous work-history sample overstates by 0.3 percent the number of workers with wage credits during 1937–50. The actual overall percentage may be a little higher, and the percentage is, of course, higher for certain groups of workers and lower for others.

Measuring sampling error.—Since the data are based on a sample, they are subject to sampling error. In the past the sampling error of various estimates has been calculated on the assumption that the sample precision is approximately that expected from a simple random sample. At present a special study of the sampling error is being made. The 1-percent sample is being split into strata for each of many characteristics, the variances within each stratum are being estimated, and these variances are being added for all groups.

Data for most of the cells are presented as single figures, and estimates of the universe can be made by adding two zeros to the sample figures. One significant feature of the conversion of the vast accounting processes to electronic processing is the possibility of obtaining about 200 totals from the 20-percent subuniverse from which the 1-percent sample is selected. These totals will provide a basis for ratio estimates that will improve the precision of various estimates derived from the sample.

Size of sample.—Though the 1-percent sample is a large sample—more than 1 million units—it is too small for certain types of program studies. The research and analytical program of the Bureau includes some projects that deal with small segments of the total coverage, such as the self-employed or household workers. When a detailed analysis is attempted of such groups with the 1-percent sample, only a few thousand accounts are available for study. If the sample for such a group is studied for several cross classifications, many cells occur with only a few workers, and the sampling error becomes large. When an attempt was made, for example, to obtain data on the number of older insured workers by county, many counties showed a small number of aged insured workers. The resulting figures were subject to large sampling error. This type of data is needed in estimating the number of potential claims for benefits in the district offices of the Social Security Administration. The problem lies in the fact that the continuous work-history sample is a flat 1-percent sample throughout the country, whereas samples of varying size are desirable for special needs. Thus far the development of samples of different size for different groups of workers has not proved feasible.

Timing.—Another problem in the work-history sample is timing. One-percent sample data cannot be obtained for and through a given year until at least 12 months after the end of that year. Data on the 1955 and 1937–55 work experience, for example, will not be available until sometime in 1957. This time-scheduling problem results primarily from the fact that the basic information for the sample flows from the Bureau's earnings processing operations. The posting operation from which data for a given calendar year are obtained is not completed until August or September of the following year. As a result, the 1-percent sample file is received 8–9 months after the period to which the data refer, and only then can the statistical and summarization operations be started. They are followed by a quality-assurance program and then by publication of the data. In the past there has been a lag of at least 1½ years before the data for a given year could be published. With the aid of electronic equipment, it is hoped that ultimately this lag will be cut several months.

One method of dealing with the problem of delay is the development of an advance sample, which is a sample of one-tenth of 1 percent, from which more nearly current data on work and wage histories can be obtained. With this small sample it has been possible to reduce the lag to about 1 year. The data cannot be as detailed, of course, as those from the 1-percent sample.

Advantages of the Digital Sample

Since the continuous work-history sample is based on digits in the nine-digit social security account number, it may be appropriately termed a "digital" sample. Such a sample has several important advantages over other sampling methods, at least for maintaining a continuous sample. Some of these advantages are simplicity, flexibility, and economy.

The method of sample selection is simple to apply mechanically. As a result, complications due to incomplete selection or overselection of units by the operators are automatically avoided.

A digital sample lends itself to easy subsampling by simply sorting on the basis of the digits in the account number. In addition, the Bureau's sample records can be compared directly with records similarly selected on the basis of digits in the social security account number by the unemployment insurance and railroad retirement agencies.

Because this type of sample can usually be efficiently integrated with other regular operations required by the agency using it, sample selection is less costly than other types. Statistical controls established in these earlier operations also help to make unnecessary a review of inconsistencies that would otherwise show up later.

Summary

The continuous work-history sample consists of a permanent and self-sustaining 1-percent sample of accounts established for old-age and survivors insurance purposes since 1936. Basic records are being accumulated for each account holder, thus providing a new type of statistics—that is, data on patterns of employment and earnings over long periods of time. Each year data are tabulated by a variety of classifications, such as age, sex, insurance status, geographic and industrial mobility. These data are used in evaluating the operation of the old-age and survivors insurance program and in administrative planning.

(Continued on page 27)

Table 5.—*Status of the unemployment trust fund, by specified period, 1936–56* [1]

[In thousands]

Period	Total assets at end of period [2]	Net total of U. S. Government securities acquired [3]	Cash balance at end of period	State accounts				Railroad unemployment insurance account [6]			
				Deposits	Interest earned	Withdrawals [4 5]	Balance at end of period	Deposits	Interest earned	Withdrawals	Balance at end of period [7]
Cumulative, January 1936–December 1956	$9,069,279	$9,061,089	$3,138	$22,366,176	$2,381,315	$16,201,186	$8,546,305	$1,246,618	$208,182	$1,137,382	$317,418
Calendar year:											
1940	1,957,977	436,300	12,677	860,784	58,901	614,814	1,804,835	59,907	1,217	15,449	153,142
1941	2,744,358	786,700	12,358	1,008,149	53,000	349,583	2,516,400	66,281	4,557	15,088	227,958
1942	3,698,008	955,000	11,008	1,138,530	68,047	344,263	3,378,714	85,973	6,084	6,695	319,293
1943	5,146,745	1,408,000	51,745	1,328,117	81,864	77,582	4,711,113	98,244	7,409	1,014	435,632
1944	6,583,434	1,484,000	4,434	1,316,940	50,518	63,153	6,015,418	119,261	4,564	568	568,016
1945	7,537,391	929,184	29,208	1,160,712	118,460	461,709	6,832,880	117,374	11,010	1,949	704,511
1946	7,585,255	55,816	21,255	915,787	130,183	1,103,967	6,774,884	122,053	13,347	39,168	810,371
1947	8,124,162	538,487	21,675	1,097,213	131,620	786,875	7,216,842	126,360	15,574	54,862	907,320
1948	8,520,442	393,878	24,077	989,067	218,902	852,484	7,572,327	67,001	27,333	60,120	948,115
1949	7,780,021	−800,068	52,125	997,582	156,472	1,761,695	6,953,683	7,133	19,190	146,241	826,338
1950	7,721,432	−57,069	24,181	1,190,397	145,687	1,341,832	6,947,935	15,420	16,916	85,178	773,497
1951	8,519,230	787,933	99,263	1,495,218	158,265	844,672	7,756,745	19,752	16,505	47,270	762,484
1952	9,032,018	595,928	16,118	1,371,660	177,351	995,549	8,310,207	20,020	16,594	77,288	721,810
1953	9,556,549	521,916	15,882	1,350,011	201,277	969,894	8,891,602	19,399	16,189	92,451	664,947
1954	8,749,444	−805,076	4,486	1,135,261	198,602	2,032,194	8,193,272	17,287	13,692	204,078	491,848
1955	8,764,415	13,614	4,692	1,214,977	184,974	1,351,551	8,241,672	16,446	9,539	145,675	372,157
1956	9,069,279	307,546	3,138	1,504,131	199,597	1,399,095	8,546,305	56,592	8,119	119,450	317,418
1955											
October–December	8,764,415	128,980	4,692	284,401	46,928	227,429	8,241,672	3,831	2,190	28,465	372,157
1956											
January–March	8,560,798	−250,000	4,113	210,131	46,882	433,468	8,065,216	3,784	2,047	33,875	344,114
April–June	8,794,426	197,938	[8] 88,294	472,710	46,724	368,763	8,215,887	19,768	1,942	20,670	345,153
July–September	8,982,117	216,980	3,582	490,906	52,627	310,105	8,449,314	16,767	2,113	32,505	331,528
October–December	9,069,279	142,629	3,138	330,384	53,365	286,759	8,546,305	16,273	2,017	32,400	317,418

[1] Beginning 1949, not strictly comparable with data for earlier years because of differences in accounting methods in source materials used.

[2] Beginning December 1954, includes assets of the Federal unemployment account, under the Employment Security Administrative Financing Act of 1954.

[3] Includes accrued interest and repayments on account of interest on bonds at time of purchase; minus figures represent primarily net total of securities redeemed.

[4] Includes transfers from State accounts to railroad unemployment insurance account amounting to $107,161,000.

[5] Includes withdrawals of $79,169,000 for disability insurance benefits.

[6] Beginning July 1947, includes temporary disability program.

[7] Includes transfers to the account from railroad unemployment insurance administration fund amounting to $97,646,000 and transfers of $12,338,000 out of the account to adjust funds available for administrative expenses because of retroactive credits taken by contributors under the Railroad Unemployment Insurance Act Amendments of 1948.

[8] Includes transfer from general funds of $81 million for the Federal unemployment account, held temporarily as undisbursed appropriation.

Source: Unpublished Treasury reports.

WORK-HISTORY SAMPLE

(*Continued from page 13*)

Major improvements are being made in the composition of the sample, beginning with the 1937–55 tabulations. These changes result from the adoption of electronic equipment for processing the sample and from the amendments to the Social Security Act that introduced new concepts and study areas. Without the electronic equipment, it would not have been feasible to maintain the work-history sample except on a substantially limited basis. With the conversion to electronic processing the potentialities of the continuous work-history sample appear bright, because of the far greater informational storage capacity, the greater speed of tabulation, and the reduction of clerical work as a result of the many complex and intricate calculations the electronic equipment can perform.

The sample is maintained on a continuing basis annually. Each year new workers are added to the sample by selecting from newly issued accounts those that contain the designated sample digits. Other maintenance operations involve identifying each year the newly retired workers, the newly disabled workers, and workers who died in the year. The year's earnings and employment experience of workers employed in the past year are added to the preceding year's accumulated record, thus bringing up to date the cumulative earnings and work records.

Many practical problems exist in developing the continuous work history to serve more fully the needs of social security analysts as well as analysts in other fields of economic and demographic research. Among these are more nearly complete identification of the accounts of deceased workers, the maintenance of fuller continuous work-history patterns, the control and effective prevention of multiple records for the same worker, improvement in the methods of estimation and measuring sampling error, developing a large sample for study of selected areas with relatively few workers, and, finally, advancing the date of completion of the tabulations or otherwise obtaining data more currently.