

The Continuous Work History Sample Under Old-Age and Survivors Insurance

By Jacob Perlman and Benjamin Mandel*

FOR THE PAST 3 years, the Bureau of Old-Age and Survivors Insurance has been developing a new approach to the study of workers covered by the social security program. This approach, designed to provide data on the continuity of earnings and employment, affords an opportunity to study patterns in the course of a person's working lifetime. The data are obtainable from a random sample of workers, whose wage records, once selected for inclusion in the sample, will remain there until their death.

Most information on earnings and employment, as collected by various agencies for many years, covers only a limited unit of time—generally a week or a single pay-roll period. Such data are useful in depicting a cross section of the earnings of workers in the labor force, but they have a distinct limitation in that they fail to trace the earnings of individuals over long periods of time. Comprehensive data on continuity of earnings and employment cannot be obtained by means of the usual statistical surveys.

Interest in continuity of earnings and employment developed considerably in the past decade. During 1900-1929, unemployment and the consequent loss of earnings was regarded as a temporary phenomenon in economic life. Though there was unemployment due to seasonal, cyclical, and other factors, it generally did not affect the same individuals over long periods. During the thirties, however, when from 8 to 16 million persons were out of work at any one time, wage losses from unemployment became a chronic phenomenon and the need for statistics on continuity of earnings and employment became urgent.

Several surveys were made in the thirties in response to the demand for information on duration of unemployment. For the most part, these surveys used the technique of individual interviews to obtain work histories going back over several years. Most of the surveys, however, were limited in coverage and were based on the memory of the persons interviewed, which cannot be relied

upon for accurate wage and employment information reflecting long periods of time.¹

Adoption of the Social Security Act in 1935 opened up the possibility of developing comprehensive and accurate data on continuity of earnings and employment. The act requires employers to make quarterly reports on taxable earnings of individual workers. Utilizing these reports, the Social Security Board developed a system of wage recordkeeping which made it possible to determine at any time the total taxable wages received by an individual from various employers in covered industry. This system also provided a basis for tracing the patterns of earnings and employment of a given worker over an appreciable period of time.

By the end of 1937, the first year the program was in operation, the Board had established ledger accounts for more than 36 million holders of social security account numbers. At some time during the first year, 32.6 million workers earned taxable wages under the program, amounting to \$29.4 billion. To record these wages to the account of each individual worker involved processing about 2 million separate employer reports for each half year, and 77 million separate wage items. By the end of 1943, the total number of accounts established had mounted to nearly 76 million. It is estimated that there are 69.4 million living social security account-number holders, about two-thirds of the total population of working age in the country.

Early Wage Tabulations of the Bureau

In 1938, the Bureau of Old-Age and Survivors Insurance began tabulating wage-record data on annual earnings in 1937 according to the age, sex, and race of the worker and the State in which he was employed. This type of annual tabulation

¹ The 1940 census of population obtained information on a comprehensive basis on duration of unemployment up to March 30 of that year, as well as the number of equivalent full-time weeks worked and income in 1939. However, this information is subject to limitations in accuracy, being based on the memory of the individual involved or some other person available for interviewing.

*Bureau of Old-Age and Survivors Insurance, Analysis Division.

was continued for 1938 and 1939; the data for these years were also tabulated by industry and by cumulative wages since 1937. The 1937 and 1938 tabulations were based on a 100-percent coverage, whereas the 1939 tabulation was confined to a 20-percent sample. These tabulations provided information annually on the coverage of the program and also filled an important gap in wage statistics on annual earnings. Except for the cumulative amount of taxable wages, however, they provide no information on the continuity of earnings and employment for the periods covered.

In August 1939, Congress amended the Social Security Act and set up continuity of employment and wages as a basis for eligibility and benefit amounts under the system, which significantly affected the type of statistics the Bureau required for appraising the operations of the program. The original act based benefits on cumulative wages earned and the number of years employed since 1936. Therefore, the Bureau could determine the number of covered workers, the number and proportion of insured persons, and the potential amount of benefits or lump-sum payments by simple tabulations showing the distribution of workers according to their cumulative wage credits. Under the 1939 amendments, on the other hand, an individual is eligible for old-age benefits upon attaining age 65, if he has had at least 1 quarter of coverage (a calendar quarter in which he has received taxable wages of \$50 or more) for every 2 calendar quarters which have elapsed since 1936, or the quarter in which he attained age 21, whichever is later.² Since his benefit amount is computed on the basis of average wages (cumulative wages divided by elapsed time) the Bureau must have information not only on the cumulative wages of a worker but also on the specific quarters and number of quarters which were quarters of coverage.

In 1940 the Bureau set up a very small sample, comprising about one-tenth of 1 percent of all accounts established during the years 1937-39. The sample was selected on the basis of specific digits in the account numbers of individuals registered in 22 States, and the resulting tabulations showed the specific years in which a worker was in covered employment, the number of his

² For the detailed provisions of the act, see "Federal Old-Age and Survivors Insurance: A Summary of the 1939 Amendments," *Social Security Bulletin*, Vol. 2, No. 12 (December 1939), pp. 3-16.

quarters of coverage and employment, his cumulative wages for the 3 years, his age, race, and sex, and his annual earnings.

These data served immediate needs for preparing administrative and actuarial estimates, and they threw some light on the reasons for lack of insured status.³ The tabulation, however, had some serious limitations. It was not representative of the various areas of the country, since the 22 States sampled did not include some important industrial States. It was too small to permit detailed analyses to test many aspects of the existing provisions of the act and suggested alternatives.

Continuous Work History Sample

For purposes of long-range analysis, therefore, the Bureau designed a much larger sample covering all parts of the country and using all significant information on covered workers. In setting up this sample and tabulating the data, special emphasis was placed on the relationship of the individual's work history to his insurance status and potential insurance benefits.⁴ The system is sufficiently flexible to make available, without much extra effort, data on earnings and employment that can prove illuminating in studying other present and post-war economic problems.

The sample is intended to include, at any one time, 4 percent of the accounts of all living workers with wage credits. The initial selection covered accounts established during the period 1937-40 and was made on the basis of certain digits in the social security number. Each year, according to the plan, accounts of "new" workers, selected on the basis of the same digits in the account number, will be added to the sample, the current year's wage record of all workers in the sample will be added to the accumulated information, and the accounts of workers who have died or become entitled to primary benefits will be segregated from the others.

Planning and codifying the information on punch cards has been a prodigious task.⁵ Because of the wartime need for economy in personnel time, only one-fourth of the sample has so far been included

³ See Wermel, Michael T., and Mandel, Benjamin, "Insured Status Under Old-Age and Survivors Insurance," *Social Security Bulletin*, Vol. 4, No. 11 (November 1941), pp. 3-7.

⁴ The sample can also be used in studying eligibility and benefits under unemployment compensation programs.

⁵ The development of details in the original plans was largely the work of Lawrence Alpern, Charles Goor, George E. Immerwahr, Ruth A. Keller, and Joseph A. Smith.

in the tabulations which are now available for 1937-40 and 1937-41 and will soon become available for 1937-42. After the war, it is hoped to include the full 4-percent sample.

Method of Selection of the Sample

To understand the method used in selecting the sample, it is essential to present some of the facts about the Board's system of issuing social security account numbers, of setting up employee wage records, and of maintaining these records.

Account numbers are issued to individuals by the field offices of the Board. Blocks of numbers are assigned these offices in a systematic order by the Accounting Operations Division, which maintains a master control of all social security numbers. Each field office must issue account numbers consecutively, starting with the lowest number of the block assigned to it.

Once issued to an individual, the account number serves as the basis for identifying all his records under the program, not only the wage records throughout his working lifetime but also his benefit records. All files used in maintaining the records are kept by account numbers in consecutive order. The first three digits of the account number represent a specific State, the next two a specific group of cards within the State, and the last four—designated as the serial number—a specific individual in any one State or group. Thus, the first filed record represents a person with account number 001-01-0001; the second, 001-01-0002; the third, 001-01-0003, and so on. Since the serial number varies for each individual within a specific State or group, the last four digits were utilized as the basis for selecting the sample.

Since the taxable wages are not posted through a complete calendar year to all accounts, the selection of the sample for statistical purposes had to be confined to a specific group of cards for which the cycle of posting provided data for a full calendar year.⁶ This group of cards, which is designated as posting group A, consists of accounts which have either a 2 or a 7 as the first digit of the serial number, and represents 20 percent of all accounts established.⁷

⁶ For description of the method of cyclical posting of taxable wages, see "Old-Age and Survivors Insurance Statistics—Employment and Wages of Covered Workers, 1939," pp. 12-14.

⁷ This 20-percent sample was used in preparing the annual tabulations for 1939.

The 20-percent sample, comprising posting group A, could have been selected on the basis of the last digit in the account number; thus, out of each 10 numbers issued, 2 would have been included in the sample. Such a method of selection, consequently, would theoretically have produced a more representative sample than the present system, which does not yield a 20-percent sample until 500 account numbers have been issued. Although this method seemed most desirable from a statistical standpoint, it was not adopted for practical reasons. Since the files for maintaining the wage records are in account-number order, a procedure of selecting the accounts on the basis of the last digit would have made it necessary either to sort out the punch cards to be included in the sample each year on the basis of the last digit, or to keep the accounts in a separate file after they were segregated the first year. The former method would have been very expensive, involving considerable sorting; the latter would have involved extensive changes in the accounting procedures, which require maintenance of all accounts in strict numerical order. Consequently, it was decided to base the 20-percent sample on the first digit of the serial number, each representing a block of 1,000 accounts. All accounts with either a 2 as the first digit of the serial number (from 2000 to 2999) or a 7 (from 7000 to 7999) could then be selected, without having to sort on the last three digits in the account number, or without establishing a separate file.

Furthermore, since the volume of accounts issued in the country, or even in the various States, is large in absolute numbers, this method of selection has produced a representative 20-percent sample.⁸

The 4-percent continuous work history sample was then obtained from this 20-percent sample by selecting accounts having 0 or 5 as the last digit of the account number. The 1-percent sample, to which the work history tabulations are confined for the time being, was in turn obtained from one-fourth of the 4-percent sample and comprises accounts with a 2 or a 7 in the sixth digit and an 05, 20, 45, 70, or 95 as the last two digits of the account number. Selection on this basis, with both high and low digits almost equally represented, not

⁸ For a discussion of the representativeness of the 20-percent sample, see "Old-Age and Survivors Insurance Statistics—Employment and Wages of Covered Workers, 1939," pp. 15-25, 350-365.

only ensured the randomness and representative-ness of the sample relative to the total universe, but also provided a basis for adding "new" accounts to the sample each year by selecting these digits in the year's new accounts.

Obtaining Data for 1937-40

After the selection of the 4-percent sample of accounts issued during 1937-40, the next task was to obtain the complete wage and employment record through 1940 for each individual represented. For 1937 and 1938, wage cards were punched, one for each reporting period of the 2 years, from the information on the ledger sheets for each worker in the sample. For 1939 and 1940, use was made of the quarterly wage cards still available from the annual employee tabulations for those years. Only in cases in which it was found that a worker had an additional account, not included in the sample, was it necessary to refer to the ledger sheets for 1939 and 1940 wage data.

The information from the ledger sheets and wage cards for 1937-40 for the entire 4-percent sample was then transferred mechanically to A-1 cards bearing the employee account number. The

absolute amount of the quarterly wages for 1937,⁹ 1938, and 1939 had been punched on the wage cards in terms of wage-interval codes before transfer to the A-1 card. Accounts with wages of \$1 or more in any 1 quarter were coded "active"; all other accounts were coded "inactive." From the wage data in the A-1 card, information was derived on the number of quarters in which employed, the number of quarters in which \$50 or more was received in specific years, and other factors. This information was next transferred, after condensation of some of the items, to A-2 cards; selected information from this card, after further condensation, was then transferred to corresponding B cards, reproduced in part from the 1940 annual employee summary card. Other items of information—such as insured status, the specific years in which employed, and total quarters of coverage, 1937-40—were also punched into the B card. These B cards were used in making the detailed 1-percent sample tabulations for 1937-40.¹⁰

⁹ The 1937 wages were reported by employers on a semiannual basis.

¹⁰ Tabulations are now being made for the 4-percent sample of 1937-41 B cards representing workers aged 60 and over in 1940. Special cards have also been prepared for tabulating other special data, such as number of employer and State changes and highest quarterly wages in a given year, which are primarily for use by other bureaus of the Board.

B cards, continuous work history					
1937-40		1937-41		1937-42	
Item	Description	Item	Description	Item	Description
1	Employee account number.	1	Employee account number.	1	Employee account number.
2	Month and year of birth.	2	Month and year of birth.	2	Month and year of birth.
3	Sex.	3	Sex.	3	Sex.
4	Color.	4	Color.	4	Color.
5	Year account number was issued.	5	Year account number was issued.	5	Year account number was issued.
6	State of employment in 1940 (State of issuance of account number if not employed in 1940).	6	State of employment in 1941 (State of employment in 1940 if not employed in 1941).	6	State of employment in 1942 (State of employment in 1941 if not employed in 1942).
7	Industry in 1940 (no code, if not employed in 1940).	7	Total amount of wages; 1937-40 and 1937-41.	7	Total amount of wages, 1937-42.
8	Number of wage items reported for each quarter of 1940.	8	Amount of wages for the year and each quarter, 1941.	8	Amount of wages, 1941.
9	Total amount of wages, 1937-40.	9	Specific years in which employed, 1937-41.	9	Amount of wages for the year and each quarter, 1942.
10	Amount of wages in each year, 1937-40.	10	Number of increment years 1937-41 (years in which \$200 or more was received).	10	Specific years in which employed, 1937-42.
11	Total amount of wages, 1937-39 (cumulative wage code).	11	Specific quarters in which employed, 1940.	11	Number of increment years 1937-42 (years in which \$200 or more was received).
12	Specific quarters in which employed, 1937-40 (specific half year for 1937).	12	Number of quarters in which employed, 1941.	12	Number of quarters of coverage (1937-39; 1937-42; 1938-42).
13	Specific quarters in which a quarter of coverage was acquired, 1937-40.	13	Specific quarters in which a quarter of coverage was acquired, 1939-41.	13	Specific quarters of coverage, 1939-42.
14	Number of increment years, 1937-40 (years in which \$200 or more was received).	14	Number of quarters of coverage (1937-30; 1937-41; 1938-41).	14	Insured status as of Jan. 1, 1942.
15	Number of quarters in which \$25 or more was received.	15	Number of quarters in which employed, 1938-41.	15	Insured status as of Jan. 1, 1943.
16	Number of multiple accounts.	16	Insured status as of Jan. 1, 1941.	16	Number of quarters in which employed, 1941.
17	Insured status under a \$25 quarter-of-coverage provision.	17	Insured status as of Jan. 1, 1942.	17	Number of quarters in which employed, 1938-42.
18	Specific years in which employed, 1937-40.	18	Benefit status (active, retired, or dead).	18	Specific quarters in which employed, 1942.
19	Number of quarters in which employed, 1938-40.	19	Quarter and year of retirement or death (except death subsequent to retirement).	19	Benefit status (active, retired, or dead).
20	Number of quarters of coverage (1937-39; 1938-40; 1937-40).	20	Quarter and year of death after retirement.	20	Quarter and year of retirement or death (except death subsequent to retirement).
21	Insured status under the provisions of the act, as of Jan. 1, 1941.	21	Amount of family monthly benefits of workers attaining age 65 and subsequently retired.	21	Year of death after retirement.
22	Month and year in which sixth, seventh, and eighth quarters of coverage were acquired, 1938-40.	22	Adjudicated age (legally determined age when retired).	22	Primary benefit at retirement.
		23	Identifying card number.	23	Number of employer and State changes, 1942.
				24	Identifying card number.

It was also necessary to develop procedures for carrying the data forward,¹¹ for adding each year a sample of newly established accounts, and for adding the current year's wage and employment data for all workers included in the sample. Ways also had to be developed to handle certain technical procedures for improving the accuracy of the data, such as (1) adjusting for erroneously and delinquently reported wages which are not included in a current year's tabulations but are carried over into the next year's tabulation, (2) handling multiple accounts of workers included in the sample, (3) identifying death and retirement cases, and (4) adjusting for corrections in reported information on age, sex, and race.

Carrying Data Forward

The most economical method of carrying the sample data forward is to transfer mechanically to a new B card selected data from the preceding year's B card and selected data from the current year's annual employee summary card. The annual employee summary card is prepared from the individual quarterly wage cards for a calendar year, which represent quarterly wage items reported by employers. Thus, last year, selected data punched in the 1941 annual employee summary card were combined with selected data in the 1937-40 B card to produce data for 1937-41. At present, selected 1942 annual employee data are being combined with selected data on the 1937-41 B card to produce a continuous work history card for 1937-42.

Thus far the continuous work history data have been restricted to a single punch card covering 80 columns. This limitation necessitated the selection of items primarily needed in answering questions concerning the program, and the exclusion of other desirable combinations of wages and employment information not so directly related. To provide a basis for tabulating such data in the future, however, provision has been made for storing the basic annual employee punch cards and the continuous work history cards for past years. Using these cards, it will be possible later to select any combination of obtainable data for various tabulations.

¹¹ The A-1 and A-2 cards are almost completely coded for 4 percent of the accounts, while the 1937-40 B cards are completely coded for 1 percent and partly coded for the remaining 3 percent of the accounts.

Adding New Workers

Provision has also been made for the addition each year of new workers to the sample by selecting account numbers from the current year's applications, on the basis of the specified digits in the account number, and for setting up a B card, one for each new account. Activity or inactivity of the account is determined by reference to the 4-percent sample of annual employee cards. Thus, if no wages have been reported for the new account included in the sample, the account is retained in an inactive file containing punch cards with basic identifying information but no wage and employment data. In later years, whenever wage data are reported for accounts in this inactive file, they are to be incorporated into the punch card, which is then included in the tabulation of active cards.

Carry-over and Other Wage Adjustments

Many employers report taxable wages too late to be included in a given year's tabulation, and some wages, even though promptly reported, cannot be identified for posting to the proper worker's account and are thus also excluded from the tabulation. Wages reported or identified too late to be included in the tabulation for a given year or years—the "carry-over" wages—are held for inclusion in a subsequent year's tabulation.

Since the continuous work history data for any year can reflect only the wages posted to individual accounts through that year, the wage data shown on some of the B cards are deficient with respect to current year's wages, and—to a smaller extent—the wages of prior years. For example, the wages shown in the 1937-41 B cards exclude approximately 3 percent of all the 1941 wages that are expected ever to be processed, and about 1 percent of each preceding year's wages. The projection of the sample from one year to the next is such, however, that it permits the inclusion of carry-over wages not previously included. Thus, the 1937-42 B cards will show 1941 annual wages and 1937-42 cumulative wages which will include most of the carry-over wages omitted from the 1937-41 B cards. Since the annual and cumulative wages shown in each year's B cards reflect all wages posted through the current year, the only adjustments which need be made in interpreting the wage data are for the carry-over wages which will be posted in subsequent years.

Since many of the informational items on the

B card, such as quarters of coverage, insured status, and so on, are derived from the wage data, adjustment of wages for carry-over would also require corresponding adjustment of informational items on the punch cards. This type of adjustment was made for only a few of the informational items in the 1937-41 B cards, to expedite the tabulations. Plans have been developed, however, for adjusting for carry-over practically all the items on the 1937-42 B card, so that a study could be made of the effects of eliminating complete carry-over adjustments in the data in future years. This adjustment will also include the wage and employment data of workers in State-chartered members of the Federal Home Loan Bank System and State bank members of the Federal Reserve System, to whom old-age and survivors insurance coverage was recently extended retroactively to 1937.

Adjustments for Multiple Accounts

Many persons hold more than one account number.¹² Hence, it was essential to adjust the sample so that the resulting data would be in terms of workers rather than accounts and the wages shown for each worker would reflect his total wages and not just wages posted to one of his accounts. Such adjustments have been made in the 1937-40 continuous work history 1-percent sample for multiple accounts discovered through January 15, 1942. By consulting the ledger sheets, which showed a cross-reference for multiple accounts, all the wages of each worker within the 1937-40 sample who had more than one account were consolidated.

It should be noted, however, that all these adjustments affect only the known duplications in issuance of account numbers. Comparisons of population and account-number data indicate that as many as a million and a half additional multiple accounts may not have been discovered. Adjustment for such accounts can be made only by estimates based on special studies; however, the resulting error apparently does not affect general results seriously.

Identification of Deceased and Retired Workers

When originally selected, the sample for 1937-40 included the records of account holders who had

¹² By the beginning of 1944, about 2.2 million multiple accounts had been discovered, exclusive of the worker's primary account.

died before the end of 1940. The Bureau had made attempts to obtain reports of deaths of covered workers, but these efforts were only partly successful.¹³ Consequently, no provision was made in the 1937-40 B card for indicating whether or not the individual had died.¹⁴ The 1937-41 B card, however, instituted an identifying code for accounts on which a survivor monthly benefit or a lump-sum death claim had been awarded. This identification was made possible through the collation of the annual claims punch card, used in tabulating claims data, and the corresponding continuous work history B card. No effort has been made to identify deaths which have not given rise to claims. The possibility of using a field study or records of the Bureau of the Census for identifying the files of deceased persons is under consideration.

Data relating to retirement claims have also been incorporated in the 1937-41 and the 1937-42 B cards, showing when the worker became entitled to retirement benefits. This identification is valuable for the following reasons: it lays the ground work for adjusting the data to represent living workers and their wages, since at present such adjustments are possible only by estimate through the use of life-table mortality rates; it permits more accurate calculation of potential benefits for the living, nonretired groups of workers; it provides the basis for special studies of the work experience of old-age beneficiaries who have returned to work after retirement; and it permits comparison of the date of birth reported on the account-number application with the date of birth established at the time a claim is awarded, and so aids in adjusting the sample for misstated ages.

Corrections in Reported Identifying Information

Changes or corrections are often made in the Bureau files in information on age, sex, race, and other characteristics initially furnished by individuals in applying for account numbers. Since such changes relate to less than one-tenth of 1 percent of the cards in the sample, no adjustments have been made. A record of such changes is being kept, however, and the necessary correc-

¹³ See the *Bulletin*, May 1940, pp. 64-65, for a description of the "proof-of-death" program.

¹⁴ The A-1 and A-2 cards contained a "proof-of-death" code which was never utilized.

tions can be made if the number of cases ever becomes significant.

A much more frequent source of change in age arises from claims records. As previously noted, the comparison of the age established in the claim adjudication with that found in the original records is instrumental in interpreting the sample figures on age distributions.

Controlling and Checking Representativeness

When the continuous sample was established, universe data for 1937-40 were developed to check its adequacy and representativeness. The first set of universe data provides a basis for determining the percentage of total cumulative wages included in the sample tabulations. The accounting system of the Bureau requires that control figures on the total amount of cumulative wages posted to employee ledger accounts shall be obtained as of the end of each annual posting cycle. Since the "cut-off" dates for both the sample tabulations and the accounting period coincide, the sample and universe figures on wages are comparable, after minor adjustments. The size of the sample, in terms of wages, may therefore be determined by dividing the total cumulative wages included in the sample by the total universe wages.

The universe figure on cumulative wages, however, does not contain a distribution by age, sex, and race of the workers. To test the adequacy of the sample in terms of these characteristics it was necessary to develop a series of universe figures from the data on applicants for account numbers. These figures were obtained by combining the regular 100-percent annual tabulations of applicants for account numbers,¹⁵ distributed by age, sex, and race, and comparing the results with the corresponding sample data. These data are kept current by adding to the cumulative total the number of new accounts established each year.

Significance of the Continuous Sample for the Social Security Program

The Social Security Board is responsible under the Social Security Act for "studying and making recommendations as to the most effective methods of providing economic security through social insurance . . ." In setting up the continuous work history sample, therefore, consideration was

¹⁵ The tabulations for 1936-37 were based on a 10-percent sample.

given to obtaining both the type of information required in administering the existing provisions and that needed in analyzing the effectiveness of these provisions and in testing various alternatives. Thus, account was taken as far as possible of the requirements of the unemployment compensation and other existing insurance programs, and proposals for extension of social insurance to cover additional risks.

For appraising the effectiveness of the program, the sample is designed to provide data for studying the eligibility provisions, the benefit formula, and the work clause. Tabulations have been and are being prepared to show the cumulative number of persons contributing to the insurance fund and the cumulative amount of such contributions; the number who are in currently or fully insured status; the potential benefit amounts of the insured groups; and the number of noninsured workers and the amount of their contributions, according to the length of their attachment to covered industry. The latter information is being obtained by tabulating the cards according to the number and specific patterns of years and quarters during which the workers were employed in covered industry. Tabulations are also being prepared of the number of retired beneficiaries who have returned to covered employment and the amount of their current quarterly earnings as compared with their benefit amounts.

The sample also provides basic time series required in preparing estimates needed in administering old-age and survivors insurance, such as the expected claims load or the volume of wage records to be processed. It can be used to ascertain the number of persons insured each year according to specific patterns of years employed and quarters of coverage; such data give a basis for estimating the number of insured persons in future years, and potential claims loads. Data on the number of persons employed in a year, according to the number of wage items reported by their employers, serve for estimating the number of wage reports to be processed and, consequently, the number of wage items to be posted to employee ledger sheets each quarter and year. Advance estimates of this kind permit a planned and more orderly administration.

For purposes of unemployment insurance, the sample is providing data for the Nation as a whole and by States on the amount of annual taxable

earnings of individuals, and on their highest quarterly wages. This information will be used in testing the effects of various eligibility provisions and benefit formulas. Plans also have been made to tabulate data on labor mobility in covered employment, measured in terms of the number of times a worker changes employers or States, to aid in determining the effect of mobility on benefit rights and benefit amounts under the various State unemployment insurance systems.

Broader Uses of the Sample

The sample also provides information for analyzing economic factors related to labor-market conditions. For example, data available from

the 1937-40 and 1937-41 tabulations confirm the fact that there is a greater movement of workers into and out of the covered labor force than originally estimated (table 1); they also show the intensity of this flow as it is affected by changes in economic conditions from year to year.

Migration of labor and labor turn-over, which are relatively unexplored aspects of the labor market, may be clarified by the sample data. Codes on the number of employer and State changes for each worker during a year provide a basis for studying the extent of labor mobility according to sex, race, and age characteristics. Such data, without age, race, or sex distributions, have been tabulated for 1939 and 1940 (see

Table 1.—Number of workers with taxable wages at some time during 1937-41 and amount of such wages, by year of employment and sex, 1-percent sample of wage records ¹

Year of employment	Total		Male		Female	
	Workers	Taxable wages	Workers	Taxable wages	Workers	Taxable wages
Total.....	518,919	\$1,567,330,190	359,271	\$1,306,095,202	159,648	\$260,334,088
1 year only.....	115,435	25,635,981	71,122	18,666,971	44,313	6,969,010
1937.....	20,567	6,671,456	18,630	4,722,450	10,028	1,948,097
1938.....	6,650	540,237	3,660	345,387	2,990	191,850
1939.....	7,846	772,540	4,704	562,870	3,142	209,670
1940.....	9,038	905,034	6,318	637,984	3,720	267,100
1941.....	62,325	10,740,655	38,801	12,398,271	23,524	4,348,384
2 years only.....	70,608	60,184,504	49,132	45,068,310	27,476	15,116,194
1937 and 1938.....	15,304	13,045,154	8,698	9,700,616	6,606	4,244,538
1937 and 1939.....	2,053	720,430	1,343	538,459	710	181,977
1937 and 1940.....	1,567	581,862	1,112	448,215	455	133,647
1937 and 1941.....	5,389	3,446,555	4,007	2,882,700	1,382	563,849
1938 and 1939.....	4,010	1,290,175	2,220	792,896	1,790	497,279
1938 and 1940.....	796	140,518	524	102,979	272	37,539
1938 and 1941.....	2,026	791,977	1,395	628,835	631	163,142
1939 and 1940.....	5,981	3,224,136	3,694	2,477,484	2,287	740,652
1939 and 1941.....	3,873	1,555,875	2,811	1,268,033	1,062	287,842
1940 and 1941.....	35,603	34,487,816	23,328	20,228,087	12,275	8,259,729
3 years only.....	61,179	87,022,416	40,416	65,692,011	20,763	21,330,405
1937, 1938, and 1939.....	13,035	21,655,539	7,346	14,610,332	5,689	7,000,207
1937, 1938, and 1940.....	1,555	1,227,209	1,019	903,629	536	323,580
1937, 1938, and 1941.....	4,035	4,746,481	2,911	3,008,694	1,094	837,787
1937, 1939, and 1940.....	1,920	1,620,460	1,330	1,253,956	581	366,504
1937, 1939, and 1941.....	1,172	885,726	930	772,722	233	113,004
1937, 1940, and 1941.....	5,527	7,970,749	4,343	6,881,393	1,184	1,089,356
1938, 1939, and 1940.....	3,327	2,677,183	1,946	1,748,123	1,381	929,060
1938, 1939, and 1941.....	1,868	1,167,861	1,293	919,388	575	248,473
1938, 1940, and 1941.....	2,283	2,156,552	1,689	1,774,202	591	382,350
1939, 1940, and 1941.....	26,457	42,914,650	17,561	32,680,572	8,896	10,034,084
4 years only.....	56,727	131,917,561	39,207	102,637,404	17,520	29,280,167
1937, 1938, 1939, and 1940.....	17,129	47,661,901	10,970	35,450,231	6,159	12,211,730
1937, 1938, 1939, and 1941.....	5,251	9,130,063	3,747	7,221,790	1,504	1,914,273
1937, 1938, 1940, and 1941.....	6,161	12,240,010	4,763	10,480,311	1,401	1,768,729
1937, 1939, 1940, and 1941.....	9,849	23,686,476	7,708	20,747,074	2,051	2,939,402
1938, 1939, 1940, and 1941.....	18,334	30,184,021	11,929	28,737,998	6,405	10,440,023
5 years.....	208,970	1,262,569,728	159,394	1,074,930,506	49,576	187,639,222
1937, 1938, 1939, 1940, and 1941.....	208,970	1,262,569,728	159,394	1,074,930,506	49,576	187,639,222

¹ Because of mechanical processes and extent of adjustments for delayed wage records, multiple accounts, etc., data may differ slightly from corresponding data in other tabulations.

accompanying list of items included on the card). In the near future, the data will also be classified by industry and hence will show the extent of the movement of workers from one industry to another. The stability of a worker's employment, measured in terms of his attachment to a given industry over a period of years and his earnings from that industry, may also be determined. Likewise, information will be available to indicate the industry and earning patterns of workers in covered employment after they have been separated from their usual industry.

<i>Special labor mobility Y card, continuous work history</i>	
1939-40	
<i>Item</i>	<i>Description</i>
1	Account number of employee.
2	1937 wage intervals, semiannually.
3	1938 wage intervals, quarterly.
4	1939 wage intervals, quarterly.
5	1940 quarterly wages.
6	Annual wages, 1937-40.
7	Cumulative wages, 1937-39.
8	Proof of death.
9	Number of multiples.
10	Number of wage items, 1939-40.
11	Number of employer changes, 1939-40.
12	Number of State changes, 1939-40.
13	First State of employment in 1939.
14	First State of employment in 1940.

State distributions should also be of particular interest for the light they throw on the employment experience of workers who moved across State lines during the war and the extent to which such migration will entail post-war readjustments. The potential measurement of the extent to which the covered labor force has been increased in various States with persons not regularly members of the labor force, such as housewives, the very young, and the very old, should also be significant in anticipating problems of post-war readjustment.

There has been considerable discussion concerning the fluctuations in the annual earnings of workers, but little information has been available except those tabulated from the 1937-40 B cards (table 2). There are, for example, certain individuals whose annual earnings continue to rise throughout most of their working life, while many others reach their maximum earnings earlier in life and then suffer a more or less gradual reduction in earnings because of lower wage rates or intermittent employment. This phenomenon is closely related to the problem of the older worker in industry. In ordinary times, technological changes

within industry and the consequent decline in skill tend to favor the employment of younger workers. Discrimination against the older worker, therefore, may be reflected in the decline in annual earnings throughout his later working years. It is hoped that future tabulations of the sample will shed new light on the pattern of an individual's taxable wage income over a working lifetime and the approximate extent to which technological displacement of labor has become a factor in industry.

Another aspect of this problem is the extent to which unemployment due to seasonality and business depressions affects the patterns of an individual worker's earnings history. It will be possible to study patterns of earnings for workers in seasonal industries. Business depressions cause more extensive unemployment, and the consequent loss of earnings will have to be studied on a much wider basis.

Still another use of the sample data may be in measuring duration of employment or unemployment. The Bureau has no information on the total number of man-hours or man-weeks worked during the quarter, and duration of employment can only be measured in terms of quarterly earnings and man-quarters. If no earnings are reported for a given quarter, the worker probably had no covered employment. If earnings are reported for a quarter, some measure of the extent of a worker's employment or unemployment in covered industry can be obtained from his quarterly wages. For example, the quarter of coverage, which is defined by the act as including taxable earnings of \$50 or more, may be used as a rough indication of the extent of employment, and it is also possible to define employment in terms of quarterly wages of some given amount above \$50. On the other hand, it is difficult to use these earnings to indicate duration of employment in terms of units of time less than a calendar quarter unless certain assumptions are made concerning previous quarterly wage levels compared with the current quarterly wages. It is hoped that certain techniques may be developed so that sample data can measure duration of employment and unemployment for limited groups of workers, such as those belonging to specific industries.

Limitations of the Sample

Since the continuous work history sample is based on records kept for the purpose of paying

old-age and survivors insurance benefits, it is necessarily limited by the provisions of the Social Security Act, particularly the restriction of the coverage of the insurance system. It is, of course, impossible to determine from the wage records the amount of a person's earnings in noncovered employment, which comprises the major areas of agricultural employment; employment by Federal, State, and local governments; domestic service in private homes; self-employment; employment covered by the Railroad Retirement Act; and employment in certain nonprofit institutions. This limitation will be lessened or overcome if coverage is extended. In the meantime, a distribution of the data by industry can minimize the effect of the limitation, since shifts between covered and noncovered employment are believed to be restricted largely to certain types of industries.

Exclusion of earnings over \$3,000 a year from the old-age and survivors insurance system constitutes another limitation. Beginning with 1940,

the sample data include only the first \$3,000 of an individual's wages in a calendar year, in accordance with the definition of "wages" under the Social Security Act for purposes of computing benefit amounts. Thus, workers whose annual earnings in covered industry are more than \$3,000 and who may have received their first \$3,000 during the first or second quarters would not be shown with earnings in the third and fourth quarters, even though they continued in employment throughout the year at a high wage level. Adjustment for this limitation may be made by segregating the continuous work history cards representing workers in this wage class and treating them separately; it would be reasonable to assume that these workers are employed in all quarters of a year.

Since the continuous work history sample is based on a random selection, some sampling errors in the tabulated data are unavoidable. The errors caused by incomplete adjustment for multiple accounts and for carry-over wages, as well as in-

Table 2.—Number of workers with taxable wages in each year 1937-40, by amount of such wages in 1937 and 1938, in 1938 and 1939, and in 1939 and 1940, 1-percent sample of wage records¹

Taxable wage interval and year	Taxable wage interval in specified year							
	Total	\$1-100	\$200-500	\$600-1,100	\$1,200-1,700	\$1,800-2,300	\$2,400-2,900	\$3,000 and over ²
1933								
1937 total.....	224,792	26,840	46,524	73,207	43,408	18,812	7,440	8,456
\$1-100.....	25,154	14,204	8,585	2,011	211	38	9	6
200-500.....	41,836	7,650	22,088	10,958	945	121	46	19
600-1,100.....	71,127	3,644	12,471	46,928	7,474	460	96	45
1,200-1,700.....	48,370	1,091	2,915	11,545	28,698	3,815	226	89
1,800-2,300.....	21,484	126	366	1,509	5,429	12,290	1,586	172
2,400-2,900.....	8,387	24	69	190	604	1,768	4,520	1,212
3,000 and over.....	8,425	11	30	60	137	305	963	6,913
1939								
1938 total.....	224,792	10,150	40,136	74,556	50,528	22,538	8,517	9,317
\$1-100.....	26,840	12,300	10,173	3,672	534	67	6	7
200-500.....	46,524	5,150	22,678	16,189	2,232	210	40	19
600-1,100.....	73,207	1,354	6,431	50,177	14,602	1,008	130	45
1,200-1,700.....	43,408	201	712	4,021	31,620	6,380	455	109
1,800-2,300.....	18,812	40	102	408	1,801	13,831	2,255	285
2,400-2,900.....	7,440	7	24	60	176	884	5,029	1,266
3,000 and over.....	8,456	8	16	20	73	142	602	7,536
1940								
1939 total.....	224,792	20,035	37,380	70,838	51,228	25,380	9,548	10,377
\$1-100.....	10,150	10,470	6,140	2,055	402	40	9	10
200-500.....	40,136	6,480	19,969	11,665	1,709	255	30	28
600-1,100.....	74,556	2,484	9,608	49,273	11,939	984	132	46
1,200-1,700.....	50,528	462	1,197	6,868	33,523	7,848	514	116
1,800-2,300.....	22,538	101	252	754	3,240	14,712	3,112	408
2,400-2,900.....	8,517	16	64	153	312	1,243	5,043	1,686
3,000 and over.....	9,317	13	51	70	94	298	708	8,083

¹ See footnote 1, table 1.

² For 1940, interval was \$3,000 even.

complete identification of deaths, can be partly overcome by adjustments, as has already been indicated.

In tabulating data based on State and industry, problems will arise from the method of determining the State of employment and industry of a worker for a given year. For purposes of the sample the worker's State of employment and industry is generally determined from the first wage card tabulated for the worker in the third quarter of the year, and all his annual taxable wages are allocated to that State or industry even though he may have earned wages in other States or industries at some time during the same year. Since some employees work in more than one State or industry in a given year, the data would understate the number of workers employed at some time during the year in a given State or industry and the wages ascribed to such workers would be overstated as far as earnings in the State or industry coded for the worker are concerned. It will be possible, however, to determine the number of workers affected by this method of selection of State or industry by codes on the number of State and industry changes made by a worker during a year.

One serious difficulty is the delay occasioned by the present system of tabulating the sample data after wage returns for all 4 quarters of a calendar year are processed and posted to ledger accounts. Data for the first calendar quarter of a given year are available about a year and 3 quarters later; data for the second quarter are not available until about a year and a half later, and so forth. As long as the data are tabulated on an annual basis, there is the unavoidable lag due to the retention of wage cards for early quar-

ters of the year until cards for the entire year are completely processed and released for tabulation. This serious lag could be avoided if a supplementary system of tabulating the quarterly wage cards on a more nearly current basis were adopted. Quarterly tabulations combining quarterly employee wages and employment data with continuous work history data are feasible, but the adoption of such a program will require the use of special punch cards and the development of new procedures. In its early stages such a project would probably require a substantial but not a disproportionate investment of effort. Once organized and systematized, however, the program may be continued at a very moderate cost, since much of the work could be done as a byproduct of accounting operations.

Conclusion

In addition to providing information on operations in old-age and survivors insurance, the continuous work history sample is adaptable to serve a variety of statistical needs. It is believed that the sample has been adequately selected and placed on a technically sound basis. Provision has also been made for controlling the sample and for adjusting the resulting data for certain inevitable inaccuracies resulting from mechanical tabulations and from the fact that the data are developed as a byproduct of wage records for social insurance purposes.

The data have already proved their present and potential value in a broad field of inquiry. As the information includes increasing periods in individual work histories, it will afford a comprehensive picture of the employment and wage patterns of this country's workers.