

THE SOCIAL SECURITY CONTINUOUS WORK HISTORY SAMPLE AS A SOURCE OF COHORT WAGE DATA

S. D. Hearn, Social Security Administration

Purpose and Background

This paper is intended as an explanation of the continuous work history sample of the old-age, survivors insurance, and disability program which is currently being administered under the Social Security Administration. More specifically, the explanation of the sample will be directed toward the use of such sample data as a source for cohort wage analysis.

Under present legislation, over nine out of ten jobs are covered under the program. The exceptions apply to work under federal civil service and under railroad retirement, incidental employment, and to self-employed doctors. Since employers are required to report periodically the taxable earnings of each of their workers, and since self-employed must report their earnings on an annual basis, a very large quantity of data is available to the Social Security Administration. From this vast accumulation of data, a sample of records is selected for statistical analysis on a continuing annual operation.

When the reports for a calendar year are received and posted to the earnings record tape files, maintained by the Social Security Administration, a 1 percent sample of tape records is set aside for statistical processing. The results of such statistical processing are available in small part from the annual handbooks which have been released by the Social Security Administration.^{1/} Many more tabulations are prepared for analyses by the statisticians, economists, and actuaries of the Administration, as well as by those of other Government agencies and public and private organizations.

Data are available to varying degrees of detail, based on size of sample. In general, the greater the size of sample, the less detail is available. For example, there are 100%, 20% 1%, and 0.1% data. The extent of detail in the varying sample sizes is shown in Appendix 1. Although the principal purpose of the continuous work history sample is for internal studies in the Social Security Administration, the sample

was designed to serve many purposes. Many other Government agencies have found the data useful in connection with their own programs as well as for general research purposes. Although a number of studies have been completed involving analyses of the economic and personal characteristics of workers employed over a period of time,^{2/} and the use of OASI data for medical research purposes,^{3/} little use has been made of the continuous work history sample for wage studies based on cohort analyses, until the past few years, when Mr. David J. Farber of the U.S. Department of Labor began his studies.

Basic Forms and Data

Cohort studies are possible, using OASI data, because of the compulsory initial registration and periodic reporting requirements under the social security system. Current wage and employment information for a worker, as reported by his employer, can be associated with the personal characteristics of the worker and the environmental characteristics of the employer.

On the worker's application form for an account number, he provides date of birth, sex, race, and other identifying information including mother's and father's name, and his place of birth. Another compulsory registration form is required to be filled out by employers when they register with the Internal Revenue Service. In this case, the employer must identify his place and nature of business. If he has more than one place of business, he is asked to provide, on a voluntary basis place and nature of industrial activity for each of his places of business. Although these data tend to remain static, provision is made for recording changes in our records.^{4/}

^{1/} The most recent handbook is: Handbook of Old-Age and Survivors Insurance Statistics, Employment, Wages, and Insurance Status of Workers in Covered Employment 1955, U.S. Department of Health, Education, and Welfare, Social Security Administration, BOASI, Baltimore 35, Maryland

^{2/} For example, see "The Continuous Work History Sample: The First 12 Years," by Jacob Perlman Social Security Bulletin, April 1951.

^{3/} For example, see "Methods of Studying the Relation of Employment and Long-term Illness-Cohort Analysis," by Thomas F. Mancuso and Elizabeth Coulter, American Journal of Public Health, November 1959.

^{4/} Periodically, large single unit employers and most multi-unit employers are contacted to determine whether changes have occurred since their previous report, in location and nature of activity at their establishments. Workers at any time may report changes in name, date of birth, or any other items previously reported on their account number application form.

In addition to these two registration forms, data are required to be reported periodically by employers and by the self-employed. In general, employers report for each employee the amount of earnings, up to the taxable limit, and the calendar quarter of those earnings. The self-employed report both their annual earnings from wages, also up to the taxable limit, and the amount of annual earnings from self-employment. Thus, no quarterly data are available from reports by the self-employed.

A third set of forms apply to the termination of the individual's working career. Benefits are payable, under the social security program, in the event of permanent disability, retirement, or death. To obtain such benefits, workers (and their families) provide information on date of termination (i.e., disability, retirement, or death), family composition, recent and estimated future earnings, military service, etc. In some disability and retirement cases the working career may be resumed, at least on a partial basis. A cohort study of such cases may be of significance to those interested in measuring impact on the labor force of previously retired and incapacitated persons.

To summarize, detailed work history records are maintained from the time the individual indicates his intention of wishing to work (by applying for a social security account number^{5/}) through his employment with each of his employers in covered employment and/or self-employment, to the end of his employment as indicated by the benefit application for disability, or retirement, or death (in which case generally, a survivor files for a benefit.)

These files are maintained in many segments to facilitate the accounting and claims operations involved. To make statistical analyses of the characteristics of workers it is necessary to collate data from the many files in order to produce a single record for a worker containing this wide variety of information.

This type of collation was done for the continuous work history sample beginning with employment in the years 1937 through 1940. Annually thereafter summarized data for individuals in the sample were added to the original data to facilitate statistical summarizations for analysis. The latest continuous work history file contains summarized data for each individual in the sample who had covered employment in any year of the twenty-four year period 1937 through

1961.^{6/} Appendix 1 shows the corresponding data which will become available in the 1937-62 summarization.

Method of Sample Selection

To select this sample a digital technique was employed. Since each worker is identified permanently by a nine digit social security account number, this number was used as the basis for selection. The first three digits of the account number indicate the State in which the number was issued. The next two digits, together with the first three digits, are used in controlling the assignment of blocks of numbers to the local social security offices which issue the individual numbers to the public. To fall in the 1% continuous work history sample, the individual's last four digits of the account number must have the following combination:

The sixth digit of the number must equal "2" or "7" (which provides the 20% sample) and the eighth and ninth digits must equal "05," "20," "45," "70," or "95."

To assure accurate representation of the sample among persons who recently obtained account numbers, for each block of 500 numbers provided by the central office to any local office, 100 represent account numbers with the digits "2" and "7" in the sixth digit position, and the remaining 400 represent other than the digits "2" and "7" in that digit position. Thus, for each issuance of 500 numbers, 100, or 20 percent represent those with the required sample digits. It is conceivable that as of the end of any one year there may not be the exact 20 percent, and consequently, 1 percent sample representation for the account numbers issued during the year. However, because of the large volume of issuances there is little likelihood of the expected sample count being significantly different from the actual sample count of numbers issued.

The 1 percent sample is further subdivided by taking an 0.1 percent sample selected on the basis of the seventh digit of the account number. Account numbers in the 1% sample with a "5" in that seventh digit position are classified as 0.1 percent sample cases. Thus, the continuous work history sample involves three sample selections: the 20 percent sample, the 1 percent sample and the 0.1 percent sample.

^{5/} Unfortunately, social security account numbers are now being used for identification purposes where employment may not be intended. E.g., identification of bank account holders by social security numbers is required by Internal Revenue Service.

^{6/} Because of processing lag, most of the earnings from self-employment in 1961 will not be included until the following year's sample file is produced. It is estimated a total of 110 million living workers are represented in this file, of which 67.6 million were employed as a wage worker in 1961.

Limitations of the Data

In considering potential uses for the continuous work history sample (and especially for cohort studies), it is important to keep in mind the limitations of the data. One of the principal limitations is the inability to pinpoint, in some cases, the reasons for gaps in an individual's work history.

In 1937, the first year in which employment and earnings had to be reported, covered employment included employment by wage and salary workers under age 65 who were engaged in non-agricultural industry and commerce. This comprised roughly half of all paid employment jobs. Since 1937, coverage under the program has been expanded, by legislative enactment, until beginning with 1955 over 90 percent of persons in paid employment include persons in covered employment.^{1/} Records of employment and earnings for the groups covered in 1951 for the first time (such as the non-agricultural self-employed) are not available in the social security files for years prior to 1951. To generalize, since only covered employment is reported, persons working in consecutive covered and non-covered employments would appear to have gaps in their employment history.

Another problem relating to gap in employment is caused by unreported deaths. Currently, about two-thirds of deaths are reported to the Social Security Administration in connection with filing for survivor benefits and/or because of reports by undertakers. Unfortunately, it is difficult to estimate the degree of underreporting of deaths in the earlier years of the program.

Another area in which impact on the data cannot be measured accurately is the multiple account number situation. Workers with more than one account number create a sampling problem because they have a better chance of having one of their account numbers fall into the sample selection. This also contributes to the gap problem because the individual's employers may be reporting his employment and taxable earnings under different account numbers. Screening procedures are used

to minimize the number of persons with multiple account numbers. A special study of the problem indicated that the number of workers in the work history sample was overstated by less than 1 percent.^{8/}

During a given year another gap in the covered employment record may appear due to the regulations for reporting earnings up to but not in excess of the taxable limit. Thus, if a worker earns his limit in the 2nd calendar quarter of a given year, his employer will not report him in the third and fourth quarter tax returns even though he may continue his employment with that employer.

Suggested Types of Cohort Studies

Given the limitations described above, at least one type of cohort study is feasible--namely, the retrospective cohort. Starting with the definition of the cohort as workers in a given industry in a recent year, an analysis could be made of changes in employment and wage patterns over a series of prior years.^{2/} The cohort could be subdivided into age-wage-sex groups to permit analyses of patterns of homogeneous groups. Annual patterns could be studied for all groups, and quarterly patterns could be analyzed for groups which do not earn the maximum taxable wages.

Another type of cohort study may be made to analyze the mobility of one or more groups of workers. For example, if we were to define the cohort as workers in covered employment in both 1957 and 1962, data could be compiled for that period on the extent of employer turnover, the frequency and distance involved in geographic change, and the extent of change in industry. Again this could be done for selected age-wage-sex groups. From a cost and time point of view, this type of study would require much more of an expenditure than the first cohort study mentioned above.

Future Prospects

With the electronic computer becoming the accepted tool for data processing, cohort studies

^{1/} There is a technical distinction in coverage between legal and reported coverage. For example, there are 8.6 million persons legally covered in State and local government and in non-profit employment, but coverage is reported for 5.8 million of that total. The remainder are workers who are eligible for coverage on an elective basis, but for whom coverage has not been arranged. The net effect of this limitation is that current social security records are available for about 85 percent of persons in paid employment.

^{8/} See The Continuous Work History Sample Under Old-Age and Survivors Insurance in the United States of America, by B. J. Mandel, First International Conference of Social Security Actuaries and Statisticians, Brussels, November 1956, page 18.

^{2/} These wage and employment data are available back to 1937 for the 0.1 percent work history sample and back to 1951 for the 1 percent sample.

should become more feasible, and subject to fewer restrictions. For example, Internal Revenue Service plans to have most of their tax return records on tape in the not too distant future. Furthermore, there is a good possibility that these tapes may be made available to the Social Security Administration. Since the same method of identification--the social security account number--will be used by both agencies, many of the limitations of the continuous work history sample would be eliminated or minimized. We would know the individual's total instead of only the taxable earnings. We would know the extent of non-covered employment and would have

better control over the multiple account and unreported death problems.

Summary

The continuous work history sample is a feasible source of cohort data, provided the limitations of the sample are taken into account. It should be possible in the future to combine for the same worker, data from tape record files of a number of government agencies such as Social Security, Internal Revenue, Census, VA, etc. When this statistician's Utopia occurs, we could truly say that just about all limitations to the data have been removed.

APPENDIX 1
EMPLOYEE STATISTICAL RECORD CONTENTS: 1962

Substantive informational contents of machinable, comprehensive, regular, multipurpose records

Item	1937-62 Workers			1962 Workers			
	All 1/			All 1/	Wage and salary		Self-employed
	Updated Summary 2/	1% CWHS	0.1% CWHS	0.1% Annual Advance Sample 3/	1% Employee- Employer 4/	0.1% Quarterly Advance Sample 5/	1% SE Sample 6/
Estimated completion date of record.....	10/63	12/63	2/64	12/63	4/64	a	5/64
Cutoff date of file (postings through).....	9/63	9/63	9/63	9/63	9/63	b	3/64
Approximate number of records (in thousands):.....	149,000	1,490	149	73	1,020	66	65
With earnings.....	131,000	1,300	131	73	1,020	66	65
Without earnings.....	18,000	190	18	n.a.	n.a.	n.a.	n.a.
Maximum record length:.....	180	350	905	80	100	80	80
Nonsubstantive items.....	30	64	70	19	7	7	38
Substantive items 7/ (listed below).....	150	286	835	61	93	73	42
<u>Account Number Information</u>							
1- Account number.....	9	9	9	9	9	9	9
2- Multiple account number indication (for multiple account numbers identified since 1946).....	X	1	X		X		
3- Year of issue of account number.....		2					
<u>Personal Characteristics</u>							
4- Year of birth.....	2	3	3	3	3	3	
5- Month of birth.....	1	2	2		1		
6- Sex.....	1	1	1	1	1	1	
7- Race.....	X	1	1		1		
<u>Insurance Status Information 8/</u>							
Insurance status, each year:							
8- 1/1/51 - 1/1/59.....			9				
9- 1/1/60 - 1/1/63.....		4	4				
Indication of insurance status based on 1956 special provisions, each year:							
10- 1/1/57 - 1/1/59.....			3				
11- 1/1/60 - 1/1/61.....		2	2				
<u>Benefit Status Information 9/</u>							
12- Benefit status, each year, 1/1/60 - 1/1/63.....		4	4				
13- Year of death for workers not previously entitled to old-age benefits or year of entitlement for workers entitled to old-age or disability benefits.....		2	2				
14- Year of death of OAB 10/.....		2	2				
15- Benefit formula code (only for living workers entitled to old-age benefits as of 1/1/63 with yr/ent. 1960-62).. 16- Indication of "in claims status".....	1	1					
Combination of family benefits of OAB's and DIB's in force status 11/:							
17- 12/31/61.....			2				
18- 12/31/62.....		2	2				
Amount of family benefits of OAB's and DIB's in force status 11/:							
19- 12/31/61.....			4				
20- 12/31/62.....		4	4				
FIA of OAB's and DIB's 11/:							
21- 12/31/61.....			3				
22- 12/31/62.....		3	3				
Indication of "in current pay status" of OAB 11/:							
23- 12/31/61.....		1					
24- 12/31/62.....		1					
<u>Disability Information</u>							
25- First year of beginning disability.....		2	2				
26- Last year of disability cessation.....		2	2				
27- First year of DIB.....		2	2				
28- Number of disability periods.....		1					
29- Indication of "freeze only" disability case and reason code.....		2					
30- Number of disability award periods.....		1					
31- Indication of disability status.....	X	X					
<u>Years Employed Information</u>							
32- First year employed.....	c 1	2	2				
33- Last year employed.....	c 1	2	2				
34- Last year employed before 1962.....		2					
Number of years employed:							
35- 1937-1962.....		2	2				
36- 1951-1962.....		2					

Item	1937-62 Workers			1962 Workers				Self-employed
	All 1/			All 1/	Wage and salary		1% SE Sample 6/	
	Updated Summary 2/	1% CWHS	0.1% CWHS	0.1% Annual Advance Sample 3/	1% Employee-Employer 4/	0.1% Quarterly Advance Sample 5/		
<u>Years Employed Information (Continued)</u>								
37- Pattern of years employed, 1957-62.....			2					
38- Indication of continuous or intermittent employment, 1937-62.....		1	1					
39- Number of increment years, 1937-50.....		2						
<u>Quarters Employed Information</u>								
Pattern of actual QE, each year:								
40- 1951-1955.....			10					
41- 1956-1960.....	X		10					
42- 1961.....	X	2	2			2		
43- 1962.....	X	2	2	2		2		
44- Actual QE with \$50 or more quarterly wages, 1962.....				1				
Deemed QE, 1962:								
45- Concept I - based on quarterly wages.....				1				
46- Concept II - based on all wages.....				1				
47- Concept III - based on all employment.....				1				
<u>Quarters of Coverage Information</u>								
Cumulative number of QC:								
48- 1937-1962.....		3	3					
49- 1947-1950.....	2							
50- 1951-1962.....	2	2	2					
51- Cumulative number of SE QC, 1951-62.....	2							
Number of QC, each year:								
52- 1937.....			1					
53- 1951-1962.....		12						
54- Number of SE QC, each year, 1951-62.....	12							
55- Number of farm QC, each year, 1955-62.....	8							
Pattern of QC, each year:								
56- 1938-1956.....			38					
57- 1957-1962.....		24	12					
Quarterly wage QC pattern 12/, each year:								
58- 1953-1960.....	X							
59- 1961-1962.....	8							
<u>Type of Work Information</u>								
60- Wage or SE indication, each year, 1951-54.....		4	4					
61- Farm or nonfarm wage indication, each year, 1955-62.....		16	16					
Farm or nonfarm SE indication, each year:								
62- 1955-1961.....		14	14					
63- 1962.....		1	1					
64- Military or other quarterly wage indication, each year 1961-62.....							2	
65- Coverage indication, 1962 (military wage, other quarterly wage, farm wage or SE).....				2				
66- Dual coverage code 1962.....					1			
67- Minister waiver form indication.....		1						
<u>Earnings Information</u>								
Cumulative:								
68- Earnings credits, 1937-62.....		6	7					
69- Earnings credits, 1951-62.....		5	5					
70- Taxable wages, 1937-50.....	8							
71- Taxable earnings, 1951-62.....	8							
Taxable wages, each year:								
72- 1937-1950.....			56					
73- 1951-1960.....			50					
74- 1961.....			5				5	
75- 1962.....			5	5	7	5		
Taxable farm wages, each year:								
76- 1955-1961.....			28					
77- 1962.....			4	5	X			
78- Taxable nonfarm wages, 1962.....				5	X			
Taxable earnings, each year:								
79- 1951-1961.....	77	55	55					
80- 1962.....	7	5	5	5				

Item	1937-62 Workers			1962 Workers			Self-employed
	All 1/			All 1/	Wage and salary		
	Updated Summary 2/	1% CWHS	0.1% CWHS	0.1% Annual Advance Sample 3/	1% Employee- Employer 4/	0.1% Quarterly Advance Sample 5/	1% SE Sample 6/
<u>Earnings Information (Continued)</u>							
Quarterly wages, each quarter - each year:							
81- 1951-1960.....			160				
82- 1961.....			16			16	
83- 1962.....			16	16	28	16	
84- Quarter in which wages of \$4797 were reached, each year, 1961-62.....						2	
85- Wages shown on SE schedule (1962) 13/.....							5
Estimated nonfarm wages, each year:							
86- 1946-1955.....			50				
87- 1956-1961 (Method 1) 14/.....			30				
88- 1956-1961 (Method 2) 15/.....			30				
89- 1962 (Method 1) 14/.....			5		7		
90- 1962 (Method 2) 15/.....			5		7		
91- Estimated earnings, each year, 1951-55.....			25				
<u>Self-Employment Information</u>							
SE taxable income, each year:							
92- 1951-61.....		44	44				6
93- (1962) 13/.....				4			
SE net earnings, each year:							
94- 1956-1960.....			30				
95- 1961.....		6	6				
96- (1962) 13/.....							6
SE industry:							
97- 1961.....		4	4				
98- (1962) 13/.....							4
SE State and county:							
99- 1961.....		4	4				
100- (1962) 13/.....							4
SE occupation (selected groups covered by 1954 amendments):							
101- 1961.....		2					
102- (1962) 13/.....							2
Indication of use of option for computing farm SEI, each year:							
103- 1959-1961.....		3					
104- (1962) 13/.....							1
105- IRD (1962) 13/.....							2
106- Type of reporting form (1962) 13/.....							1
107- Reporting year 13/.....							2
<u>Employer Information</u>							
108- Employer identification number.....					9		
109- Establishment number.....					4		
110- State and county code.....					4		
111- Industry code.....					4		
112- Class code.....					1		
113- Type of reporting form, 1962.....					1		
114- Coverage group code 1962.....					1		
115- Number of wage items, 1962.....					2	2	
116- Number of wage items, each quarter, 1962.....						8	
117- Collector's stamp, 1962.....					1		
118- Referendum code for State and local governments.....					1		

Footnotes

- a - Quarterly advance sample completion dates:
 1st quarter - 11/62 3rd quarter - 5/63
 2nd quarter - 2/63 4th quarter - 8/63
- b - Quarterly advance sample cutoff dates:
 1st quarter - 9/62 3rd quarter - 3/63
 2nd quarter - 12/62 4th quarter - 6/63
- c - For years 1937-50, record shows only that year was sometime during that period.
- 1/ Only about 75% of 1962 SE will be included.
- 2/ Number of records shown is for entire file. A 20% sample is tabulated for statistical purposes.
- 3/ Punch card file.
- 4/ Includes quarterly and annual wage workers only.
- 5/ Punch card file. Only for quarterly reported workers. File contains quarterly data for 1961 and 1962; when 1st quarter 1963 data is added, 1961 data is dropped.

Footnotes (Continued)

- 6/ Punch card file. Delinquent items in this file will be used to revise previous years' data.
- 7/ Certain fields contain more than one item of information. An "X" indicates that the item is coded in some other field.
- 8/ Insurance status codes relate to law in effect on given date:
 1/1/51 - 1/1/60 = 1:2 law
 1/1/61 = 1:3 law
 1/1/62 - 1/1/63 = 1:4 law
- 9/ Benefit status information relates only to claims based on the account number in the record. Auxiliary information is not reflected in the benefit status codes.
- 10/ Individual years 1958-1962 are shown. Prior to 1958 there is only an indication of 1957 or prior.
- 11/ Information for 12/31/62 based on family benefit record of beneficiaries in force status 12/31/62 for living workers entitled to old-age or disability benefits as of 1/1/63. Information for 12/31/61 carried over from previous year's CWSHS, with no revision for delinquently reported claims actions.
- 12/ For years prior to 1962, for each quarter that is not a QC, there is also an indication of whether earnings were:
 \$0
 <\$50 or a quarter in which there were debit earnings and a credit has been applied (does not show whether any earnings remain)
 For 1962, for each quarter that is not a QC, there is also an indication of whether earnings were:
 \$0 20-29
 01-09 30-39
 10-19 40-49
 Revision for delinquent items not complete for years 1953-55.
- 13/ Year will be other than 1962, if delinquent item in 1962 SE Sample. Informational items for delinquent reports in the SE sample refer to the reporting year coded in the record.
- 14/ Estimated nonfarm wages - Method 1:

For years 1956-58		For years 1959-62	
Quarter taxable limit reached	Estimated wages assigned	Quarter taxable limit reached	Estimated wages assigned
1st	\$16,800	1st	\$19,200
2nd	12,600	2nd	14,400
3rd	7,000	3rd	8,000
4th	4,900	4th	5,600
- 15/ Estimated nonfarm wages - Method 2:
 Consider as the "limit quarter" the one in which the taxable limit is reached. Determine the wages in the next prior quarter that shows as much or more wages than the "limit quarter." Substitute those wages for the "limit quarter" and all subsequent quarters in the year. If there is no prior quarter with wages greater than the "limit quarter," use the "limit quarter" wages for each following quarter. The total of the quarterly wages thus computed represents the annual estimated wages.

Department of Health, Education, and Welfare
 Social Security Administration
 Division of Research and Statistics
 Statistics Branch
 Statistical Processing Methods Section
 June 7, 1963

Revised June 24, 1963