The Monthly OASDI One-Percent Sample File

by Lewis F. Frain*

The Social Security Administration (SSA) compiles a 1-percent microdata sample file each month from its principal data file, the Master Beneficiary Record, which is used to administer the Old-Age, Survivors, and Disability Insurance (OASDI) program. This 1-percent OASDI sample provides current program and demographic information so that data can be quickly tabulated to respond to specific research, legislative, and administration questions about the OASDI program. This article briefly describes the development of SSA's Master Beneficiary Record and documents the contents, technical features, and uses of the 1-percent file that is developed from the master record.

*Division of Statistics Analysis, Office of Research and Statistics, Office of Policy, Social Security Administration.

In April 1989, 38.8 million beneficiaries received monthly benefit payments totaling \$18.8 billion under the Old-Age, Survivors, and Disability Insurance (OASDI) program administered by the Social Security Administration (SSA). To accomplish this task SSA relies on its principal data file the Master Beneficiary Record (MBR). From this master record of program, computational, and demographic information, SSA's Office of Research and Statistics (ORS) has designed and implemented a 1-percent monthly sample that enhances the production of descriptive statistics, simulations of the beneficiary population, and other statistical analysis and research projects. The prototype file was developed in the fall of 1984, and the first production tape was created for January 1985. A 1-percent sample file has been generated for each month since January 1985.

Development of the Master Record

When benefits for retirement and survivors insurance first became payable in January 1940, SSA established a ''payment file'' to facilitate the Social Security benefit distributions. Workers' reported earnings—the basis for determining program eligibility and the amount of benefit—were recorded in punch card form by SSA staff for geographically dispersed centers where the claims for Social Security benefits are maintained. At these centers, staffs review, allow, or disallow the claims and also certify the payment of benefits. ¹ In 1960, the payment file was converted from punch card to magnetic tape.

The first master record for the Social Security program was established in January 1962. Separate files continued to be maintained on magnetic tape at each Program Service Center. Initially, the files related only to the OASI program; information concerning the Disability Insurance program was added in July 1964. In September 1972, the Program Service Center files were combined into a single master file for OASDI beneficiaries. This Master Beneficiary Record was the first appearance of the SSA administrative record system as it looks today. So that records could be randomly accessed, the MBR was converted from magnetic tape to mass storage in 1984.

The MBR file is organized into 20 segments that are determined by

the last two digits of the claim account number (CAN), which is the Social Security number of the worker under whose account the claim for OASDI benefits was made. For example, one segment has all numbers that end in 00 through 04, another segment has all numbers that end in 05 through 09, and so on.

For each claim account number, there is one record on the MBR file. These records are of variable length and contain data on all types of Social Security beneficiaries (workers, wives, husbands, widows, widowers, children, and other dependents) receiving benefits under the respective CAN. Since October 1977, all persons who apply for OASDI benefits, as well as those actually receiving payments, have been included in the MBR file.

The file contains information about all accounts. Its primary purpose is to provide data used to calculate benefit amounts and to generate benefit checks. A tape extracted from the MBR containing the information to generate the benefit checks is sent to the Department of the Treasury in December of each year. Each month for the next 11 months only the file changes that will affect benefit payments are sent to the Treasury Department.

Before the early 1980's, the SSA Office of Operations had generated

¹ Eight Program Service Centers perform claims review and certification functions: New York, New York; Philadelphia, Pennsylvania; Birmingham, Alabama; Chicago, Illinois; Richmond, California; Kansas City, Missouri; and Baltimore, Maryland that has two centers for Disability Insurance and foreign claims.

all descriptive statistics on the OASI and DI programs. These statistics, aggregated from a 100-percent file extracted from the MBR, were produced as custom tables with fixed formats. However, because the computer system that generated these statistics was continually being modified, it lacked flexibility and timeliness.

In the early 1980's, ORS began to decentralize the production of statistics. This effort involved the creation of special microdata extracts from the MBR (both 10-percent and 100-percent samples). These new files were sufficiently abbreviated to fit on two or three reels of magnetic tape. The decentralization of the production of the statistical files was facilitated by the installation of office automation equipment that afforded ORS staff easy access to the mainframe computer. An effort was then made to keep as much as possible the generated statistics in machine readable form.

The ORS began compiling OASDI data internally in 1984 from an MBR-based 1-percent microdata sample file. Initially, this 1-percent OASDI sample was designed to produce, in a more timely manner, statistics similar to those in existing data compilations. The Office of Operations produces extracts of the MBR file quite rapidly so ORS is able to tabulate data from a file that is at most 1 month old. Direct access to microdata files also provides ORS statisticians more flexibility in compiling data to respond to specific research. legislative, and administrative questions.

Because of large sampling errors, it is not appropiate to use the 1-percent OASDI sample for studying narrowly defined subsets of beneficiaries, such as persons residing in small geographic areas.

File Organization

The process of generating the 1-percent OASDI monthly file begins with obtaining a digital sample using the claim account number from the Master Beneficiary Record. This file is provided by the Office of Operations at the beginning of each month. For each selected CAN, the extracted record from the MBR contains 552 characters of information concerning the beneficiary. The types of data contained in the file are:

Claim account number Beneficiary's own Social Security number Date of birth Sex Race of wage earner State, county, and ZIP Code Date of death of wage earner Type of benefit (beneficiary identification code, type of claim) Date of entitlement Payment status of beneficiary (ledger account file) Primary insurance amount Monthly benefit credited Benefit computation data Medicare status Supplemental Security Income status Representative payment data Dual entitlement data **Disability data** Direct deposit data

The program and demographic data provided for the sample are most valuable to the Office of Research and Statistics. Various dates of employment, earnings amounts, and administrative codes that SSA uses to determine OASDI eligibility and the benefit amount are captured. The actual data cells retained have also undergone some changes since January 1985 because it was found that some of the data are no longer useful and that other items are required.

Appropriate indicators were developed to generate OASI and DI award, conversion (from one trust fund to another), and program termination data from these monthly files. These indicators are associated with sets of conditions that occur when administrative codes from 1 month's MBR are compared with those of the previous month's MBR. This procedure involves matching records for individuals from month to month on a continuous basis. Because several persons can be selected on the same claim account number (for example, husband, wife, children) a special variable was designed to link such individuals. This variable is made up of 17 characters of information consisting of (1) claim account number, (2) date of birth, (3) sex code, and (4) beneficiary identification code.

The monthly file that is received from the Office of Operations is sorted on this special variable. The program that sorts the file also eliminates some extraneous records, and edits some codes that are difficult to process on the mainframe computers used for the statistical system. After the sort, the sample currently contains records for about 600,000 persons. These records include cases representing 38.5 million beneficiaries in current payment status, 3.3 million benefit terminations, and 1 million beneficiaries in suspended payment status. The sample selection process rejects terminated status records if the date of benefit termination is more than 12 months before the selection month.

The population of this sorted file is then matched against the 1-percent OASDI sample from the previous month using the

17-character variable described above. A determination is made using several OASDI program administrative codes, whether or not any change has occurred in the file for each individual. Each month. approximately 99 percent of the records match and no change has occurred in the program status of the individual on the file. Each month about 3,500 new awards appear on the file, and 2.500 beneficiaries have some type of change to their program status. These persons have a transaction code and a date of occurrence (month and year) appended to their record by the 1-percent OASDI matching run. The codes for major program status groups are:

Awards:

New (01-08) Converted from DI to OASI (21-19) Converted from OASI to DI (31-36) Changes in status within DI program (41-45) Changes in status within OASI program (51-67) Terminations (71-88) Conversions from one type of benefit to another (91-99, 9A, 9B)

The 1-percent OASDI sample file includes all records whether or not a transaction was made. The records with the transaction code for the current month—that is, new awards or some type of change in beneficiary status, for example, wife to widow—are also written out to a separate file called the "transaction file." This file is maintained on a tape that is stacked with similar files from other months and is used to generate award data for an entire year.

This separate transaction file is saved because only the most recent addition, change, or deletion to the master record appears on the 1-percent OASDI sample file. Any transaction that occurs in a given month will cause the previous transaction code and date field to be replaced by the current code and date. Each month an increasing percentage of records on the 1-percent OASDI sample reflect a transaction code and a date. The August 1988 file had transaction codes and dates for approximately 25 percent of the cases. The 1-percent OASDI sample process is completed at the end of this matching run.

A third program is run against the newly completed 1-percent OASDI sample to generate control tables. These tables are used to check the validity of the file. The counts from the 1-percent OASDI sample are compared with the results of the 1A Supplement Run (a monthly report summarizing program operations from a 100-percent file that is prepared by the Office of Operations) to determine if the totals agree. Table 1 is a control table generated from the 1-percent OASDI file.

The table shows the average ""monthly benefit credited" and average "primary insurance amount" corresponding to all of the groups for which population estimates are provided. The monthly benefit credited for a beneficiary reflects the actual benefit amount received and any amount deducted for Part B (Supplementary Medical Insurance) coverage under the Medicare program. The primary insurance amount is a computed monthly amount based on the worker's career earnings. The monthly benefit credited can vary from the primary insurance amount because of age at retirement, family maximum amount, and the type of

benefit. Generally, the monthly benefit credited is less than or equal to the primary insurance amount.

Ongoing and Special Uses

Since the Social Security program began, the primary use of the information gathered and recorded by SSA was and still is for administration of the program—that is, get the benefit check to the right person in the right amount in an efficient and timely manner.

However, in response to a changing society, population, and economy the program has changed and more information is needed to determine what program changes-whether by law or regulation-are to be made or anticipated. Much of the information needed to make these decisions is contained in the Master Beneficiary Record. As a byproduct of the information recorded in the MBR, sample files (such as the 1-percent OASDI sample) are drawn for use by Social Security program administrators, analysts, and policymakers. These persons have specific uses for the file: It is a data source for describing demographic characteristics of the beneficiary population, analysis of benefit offset provisions. monitoring the direct deposit program, and evaluating regional administration of claims processing. The file is used to study the longitudinal aspects of the OASDI program and to simulate the effects of proposed program changes on the beneficiary population.

From either the 1-percent OASDI sample file or from one of its auxiliary files, described below,

Table 1.-Control table for 1-percent OASDI sample: Benefits in current-payment status, August 1988

	Population	2 standard	Average primary	2 standard	Average monthly	2 standard
Total	38,413,600	73,700	\$539.06	\$0.69	\$464.66	\$0.69
Betirement	27.274.400	76,800	539.54	.86	483.01	.83
Retired workers	23,756,700	75,500	524.98	.92	515.80	.86
Men	12,462,200	62,800	628.62	1.16	581.00	1.19
Women	11,294,500	60,500	410.63	1.10	443.86	1.10
Age 65 or older	21,193,700	73,800	526.13	.96	525.06	.92
Men	11,100,400	60,100	626.35	1.23	587.27	1.28
Women	10,093,300	57,900	415.91	1.16	456.64	1.18
Age 62-64	2,563,000	31,300	515.51	2.89	439.23	2.19
Men	1,361,800	23,100	647.10	3.27	529.90	2.71
Women	1,201,200	21,700	366.31	3.24	336.45	2.41
Wives and husbands	3,076,600	34,200	647.73	2.41	267.40	1.21
Age 65 or older	2,526,500	31,100	648.49	2.72	274.58	1.37
Age 62-64	464,200	13,600	656.42	5.44	246.44	2.31
Under age 62	85,900	5,900	578.53	13.89	169.29	6.60
Wives	3,046,300	34,000	650.40	2.41	268.32	1.22
Age 65 or older	2,497,200	30,900	651.68	2.72	275.76	1.37
Age 62-64	463,200	13,600	656.78	5.44	246.56	2.31
Under age 62	85,900	5,900	578.53	13.89	169.29	6.60
Husbands	30,300	3,500	379.83	20.55	174.49	10.02
Age 65 or older	29,300	3,400	376.02	20.67	174.03	10.23
Age 62-64	1,000	600	491.52	124.3	188.00	46.49
Children	441,100	13,200	568.78	6.21	221.03	3.51
Under age 18	257,600	10,100	560.54	7.93	196.16	4.48
Students	8,300	1,800	580.46	38.82	260.65	21.92
Disabled	175,200	8,400	580.34	10.24	255.73	5.35
Disability	4,038,900	38,800	529.06	1.84	395.12	2.36
Disabled workers	2,795,900	32,600	511.54	2.24	507.28	2.26
Men	1,850,300	26,800	567.78	2.64	562.33	2.68
Women	945,600	19,300	401.50	3.07	399.56	3.09
Wives and husbands	285,500	10,700	601.75	6.31	134.23	3.17
Age 65 or older	32,400	3,600	622.92	17.06	169.03	11.10
Age 62-64	42,800	4,100	650.91	13.48	173.71	9.18
Under age 62	210,300	9,200	588.49	7.59	120.83	3.29
Wives	278,600	10,500	605.25	6.35	135.34	3.22
Age 65 or older	31,200	3,500	633.54	16.43	172.49	11.28
Age 62-64	42,700	4,100	651.16	13.50	173.80	9.20
Under age 62	204,700	9,000	591.36	7.68	121.65	3.35
Husbands	6,900	1,700	460.72	35.58	89.56	13.38
Age 65 or older	1,200	700	346.81	56.25	79.15	31.14
Under age 62	5,600	1,500	483.60	39.13	90.96	14.92
Children	957,500	19,400	558.55	3.57	145.41	2.14
Under age 18	909,200	18,900	559.36	3.67	142.01	2.18
Students	13,600	2,300	554.60	26.90	201.46	18.72
Disabled	34,700	3,700	538.81	17.56	212.56	11.13
Survivor	7,084,400	50,000	543.79	1.36	434.38	1.27
Widows and widowers	4,878,800	42,300	543.37	1.53	471.47	1.46
Age 65 or older	4,232,500	39,700	534.09	1.62	475.59	1.59
Age 60-64	646,300	16,000	604.17	4.26	444.49	3.52
Widows	4,845,600	42,200	544.18	1.53	472.41	1.46
Age 65 or older	4,211,900	39,600	534.75	1.62	476.33	1.59
Age 60-64	633,700	15,800	606.91	4.27	446.41	3.54
Widowers	33,200	3,600	424.59	18.71	334.25	16.84
Age 65 or older	20,600	2,900	399.12	22.70	325.84	21.99
Age 60-64	12,600	2,200	466.22	31.03	347.99	25.81
Disabled widows(ers)	104,400	6,500	570.04	10.26	333.82	8.83
Disabled widows	103,300	6,400	571.64	10.28	334.96	8.86
Disabled widowers	1,100	700	420.28	89.92	226.36	82.52
Widowed mothers and						
fathers	312,500	11,200	592.99	7.27	353.34	5.81
Widowed mothers	297,000	10,900	601.72	7.37	360.62	5.89
Widowed fathers	15,500	2,500	425.78	28.35	213.81	21.09

Type of benefit	Population estimate	2 standard errors	Average primary insurance amount	2 standard errors	Average monthly benefit credited	2 standard errors
Survivorcont.						
Children	1,782,000	26.300	534.94	3.11	353.15	2.34
Under age 18	1,380,800	23,200	554.35	3.61	355.62	2.78
Students	32,400	3,600	565.74	22.62	399.14	17.34
Disabled	368,800	12,100	459.53	5.55	339.86	4.09
Parents	6,700	1,600	501.79	41.37	378.36	33.67
Age 65 or older	6,400	1,600	488.98	40.08	381.08	34.40
Special age-72	15,900	2.500	146.10	.00	145.15	1.40
Primary	15,900	2,500	146.10	.00	145.15	1.40

Table 1.—Control table for 1-percent OASDI sample: Benefits in current-payment status, August 1988—Continued

statistical tables are produced for publication in the Annual Statistical Supplement to the Social Security Bulletin. These recurring tabulations are compiled for topical areas including OASDI awards and terminations, medical disability diagnoses, and representative payment.

The 1-percent OASDI sample has enabled the Office of **Research and Statistics to** respond—with data that are at most 1 month old-to questions concerning the Social Security program. Currently, ORS is examining the possibility of making an annual microdata file available for public use. The content of the file will have to be reworked extensively to ensure the confidentiality of the beneficiaries involved. The purpose of the current examination is to determine if disclosure avoidance techniques can be introduced that would preserve the usefulness of the data without compromising confidential information.

Auxiliary Files

Two files have been developed from the 1-percent OASDI sample. The transaction file, described above, is a series of monthly transaction files summarized at the end of each quarter and each calendar year to generate a set of tabulations for that period. These tabulations give descriptive statistics for program awards, conversions, and terminations for that period. Table 2 provides control data for benefits awarded in 1987.

The second file derived from the 1-percent OASDI sample is the 1-percent Disability Insurance program sample. This annual file consists of the subset of disabled beneficiaries in the 1-percent sample for December of each respective year. Because the diagnosis of the beneficiary's disability was not recorded in the MBR before July 1982, diagnostic codes were missing in a sizable number of records when they were selected for inclusion in the first annual DI program sample. In order to have a code for as many cases as possible, the initial extract was matched against various disability determination files, some of which involved coding schemes inconsistent with SSA's current diagnostic codes and required application of a translation algorithm to establish uniformity.

In creating the DI program files for subsequent years, all newly selected cases have a diagnostic code on the MBR and hence on the 1-percent OASDI sample. Older cases without a code present in the current-year file at the time of extraction will, however, have a diagnostic code on the previous DI program sample if one was obtainable from the disability determination files. Therefore, the previous and current year disability files are matched to secure codes available in the former and missing in the latter. Currently, about 83 percent of the DI cases on the file have a valid code. This proportion is higher if only the subset of disabled-worker cases on the file is examined; the other subsets. disabled widows and children. tend to have earlier entitlement dates and hence a larger portion of these records have missing diagnostic codes. Moreover, the proportion of cases with a diagnostic code continues to increase as persons leaving the program are replaced with new beneficiaries for whom a valid diagnostic code is on the MBR file.

The most recent file derived from the 1-percent OASDI sample is the "retired-worker award" file. This file was created from all of the new retired-worker awards for January 1988. The records for these newly retired beneficiaries were obtained from the 1-percent

OASDI sample, and 216 characters of data useful in examining the retirement cohorts were extracted to initiate the retired-worker award file. After this segment was built, a series of history segments was added. Each segment represents 1 month, the first being January 1988. This first history segment was filled when the file was created, and the remaining segments were initially blank filled. These history segments contain dates, benefit amounts, and administrative codes that identify the number of months for which benefits were withheld for earnings or other reasons.

For each successive month since January 1988, the retiredworker award file is matched against the current 1-percent OASDI sample. This match updates the retired-worker award file in two ways. A new retirement award for the current month causes a new retired-worker award record to be generated with the history segment for the current month. If the retired-worker award record matches the 1-percent OASDI record, the history field of the retired-worker award record is built for the current month and the demographic segment is extracted from the 1-percent OASDI record. This updated file provides a longitudinal look at retired-worker beneficiaries.

Note on Sampling Variability

The award and current benefit payment data for OASDI beneficiaries are based on a 1-percent sample taken from SSA's Master Beneficiary Record. Because the results are from a sample, they are subject to sampling variability and may differ from figures that would have been obtained had the entire MBR been used.

The reliability of data, based on a sample, depends on the degree to which sampling variability has influenced the estimates. Reliability of the data is determined by using the standard error, a statistical measure of sampling variability. The standard error of an estimate depends on the design elements such as the method of sampling, sample size, and the estimation process. About 95 percent of all possible probability samples selected with the same specifications will give estimates that produce a range that falls within two standard errors of the figure obtained from a compilation of all records.

The tables presented in this article provide an approximation of two standard errors for the estimates. For example, it is estimated that the total number of retired workers aged 62-64 is 2,563,000 (table 1). The estimated two standard errors for this figure is 31,300, which produces a range of 2,531,700 to 2,594,30.

Therefore, it can be stated with 95 percent confidence that 2,531,700 to 2,594,300 retired workers aged 62-64 receive benefit payments. As shown in table 1, the estimate for the total number of OASDI beneficiaries is 38,413,600. The estimate for two standard errors associated with this estimated count is 73.700. This estimate compares with 38,474,746 persons that the 1A Supplement Run shows as the actual number of beneficiaries in August 1988. The difference between these figures, 61,146, is less than two standard errors.

Special attention must be given to estimates based on beneficiaries residing in small geographic areas and to estimates based on other small subsets of the population, so that the associated standard errors of the variables being studied remain relatively small.

Table 2.---Control table for 1-percent OASDI sample: Benefits awarded in 1987

T						r
	Population	2 standard	Average primary	2 standard	Average monthly	2 standard
Total		errors		errors	Denetit credited	errors
	3,605,500	38,000	\$555.63	\$2.24	\$412.61	\$2.32
Retirement	2,075,500	28,800	555.76	3.09	436.45	2.94
Men	969.800	19,700	644.11	3.88	576.20	3.13
Women	680,900	16,500	391.39	4.42	361.96	3.89
Age 65 or older	485,300	13,900	563.62	6.01	566.99	6.03
Men	319,800	11,300	628.80	6.79	629.76	6.87
women	165,500	8,100	437.66	8.97	445.69	9.09
Age 62-64 Men	1,105,400	21,600	529.97	4.27 4.71	454.87	3.48
Women	515,400	14,400	376.53	5.01	335.08	3.96
Wives and husbands	306,500	11,100	638.05	7.16	245.65	3.62
Age 65 or older	66,000	5,100	606.61	17.59	266.59	9.98
Age 62-64	218,700	9,400	651.33	8.03	245.91	3.73
	21,800	3,000	599.90	26.48	179.67	13.29
Age 65 or older	295,700	10,900	646.67 635.10	7.09	248.94	3.64
Age 62-64	215,900	9,300	654.48	7.99	247.46	3.72
Under age 62	21,800	3,000	599.90	26.48	179.67	13.29
Husbands	10,800	2,100	401.89	37.94	155.56	18.02
Age 65 or older	8,000	1,800	399.40	44.61	165.67	22.00
Age 62-64	2,800	1,100	409.00	71.82	126.65	26.81
	67 700	5 200	566.02	11.00	213.02	7.10
Students	30,800	3,500	587.12	21.51	201.00	13.10
Disabled	19,800	2,800	523.29	34.11	201.37	20.16
Disability	747.500	17.300	538.61	4.50	336.75	5.82
Disabled workers	409,600	12,800	520.47	6.22	505.96	6.46
Men	265,900	10,300	588.96	7.30	573.18	7.73
Women	143,700	7,600	393.74	7.91	381.60	8.24
Wives and husbands	73,500	5,400	610.85	12.65	132.05	6.80
Age 55 of older	4,900	1,400	618.73	48.51	158.74	34.29
Under age 62	20,000 42,600	3,200	584 44	10.33	1/4.01	6 79
Wives	71,600	5,400	617.38	12.44	134.17	6.86
Age 65 or older	4,900	1,400	618.73	48.51	158.74	34.29
Age 62-64	25,600	. 3,200	656.10	17.96	175.88	12.41
	41,100	4,100	593.10	17.21	105.27	6.87
	1,900	900	364.99	81.63	52.05	31.11
Children	264 400	10 300	546.63	02.52	40.55	4.35
Under age 18	206,800	9,100	538.42	8.37	113.13	4.48
Students	51,900	4,600	580.90	15.40	200.14	10.62
Disabled	5,700	1,500	532.38	42.71	173.15	32.29
Survivor	782,400	17,700	572.53	4.53	421.88	4.38
Widows and widowers	403,800	12,700	582.84	5.96	493.27	5.96
Age 65 or older	201,100	9,000	571.44	8.84	533.52	9.45
Age 60-64	202,700	9,000	594.16	7.98	453.34	6.82
	390,200	12,500	590.41	5.90	500.75	5.97
Age 60-64	193,100	8,800	604.67	7 94	461 25	5.42
Widowers	13,600	2,300	365.68	25.42	278.32	22.14
Age 65 or older	4,000	1,300	324.70	43.74	240.17	37.26
Age 60-64	9,600	2,000	382.76	30.39	294.21	26.58
Disabled widows(ers)	15,600	2,500	592.10	24.92	331.80	24.69
Widowed mothers and	15,400	2,500	592.24	25.24	332.42	24.87
fathers	59 800	4 900	581 77	17 44	224 62	12 /6
Widowed mothers	55,100	4.700	597.00	17.79	332.36	14.03
Widowed fathers	4,700	1,400	403.19	52.83	233.95	38.66
Children	302,700	11,000	556.07	7.72	350.56	6.11
	182,100	8,500	541.28	10.23	317.67	7.80
Students	109,000	6,600	586.71	12.07	407.39	9.38
Disabled	11,600	2,200	500.16	37.68	332.86	32.58