ACTUAL SAMPLING ERROR OF SELECTED CONTINUOUS WORK-HISTORY SAMPLE SYSTEM DATA: 1972

Harold A. Grossman, Social Security Administration

INTRODUCTION. Social Security today is an extensive program, covering over 90 percent of paid employment. The major groups excluded are: Federal civilian employees; State and local government groups who have not exercised their option to be covered; and self-employed people whose net earnings (or gross income under optional provisions) do not meet the minimum reportable.

The Social Security Administration, in order to effectively administer the old-age, survivors, disability and health insurance program, requires data on the employment, earnings, and demographic characteristics of millions of workers.

The timely and economical analysis of this information depends upon the use of samples. The oldest and largest group of samples used by the Social Security Administration is referred to as the Continuous Work History Sample (CWHS) System. This is a multi-purpose system of samples of the people to whom social security numbers have been issued. For this paper we will be concerned with those who have also been reported with earnings under the program for 1972.

Information from the CWHS system is used for program purposes and can be related to the working population in general, recognizing limitations such as extent of coverage and the maximum taxable. Program uses include: research; the evaluation of program provisions such as those relating to the maximum earnings base, contribution rates, the requirements for insured status, and the "retirement test." Information can be obtained for example: on the earnings of reported workers; on the number of workers insured for different types of benefits at different ages and at different times; on the movement of workers between industries and States; and on the characteristics of people who have been reported with earnings but are not eligible for benefits.

The sample information can be prepared for a short time period, such as a calendar quarter, or for a specific year, or for more lengthy periods such as those starting at the beginning of the program, yielding information on the individual's entire experience under the program.

Speaking generally, sample data are, of course, subject to sampling error. Publications presenting data obtained from sample surveys usually have a section devoted to sampling variability. Typically, the section will begin with a statement like the following: "Since the estimates are based on sample data, they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same procedures...." The section will then continue with an explanation of the standard error, followed by examples illustrating its use. Mean square error may be covered.

In the Social Security Administration much the same procedure is followed when sample data are used. However, in addition, comparisons may be made for some items between the results obtained from sample data and from a complete census. This paper is concerned with one example—a comparison of population and 1 percent sample data on a national basis for 1972 for the characteristics: Numbers of workers, taxable earnings, and mean taxable earnings by age, sex, race and earnings interval.

Operation of Social Security Program. Before describing the population sampled and the continuous work history sample, I'll describe in a very general way part of the operation of the Social Security program. Individuals who expect to receive earnings in covered employment apply for a social security number, providing information on age, sex, and race. Social security numbers are also issued to other people. For example: children may be issued a number; applicants for and recipients of any federally funded benefits must be issued a number.

The social security number has nine digits. The first three, reading from left to right, indicate the geographic area in which the number was issued, with some exceptions. The fourth and fifth digits reflect a group within the area. The sixth through ninth places, the serial portion, are issued within each area-group to a randomized list of applicants. The serial portion is of particular importance as it is used in the sampling operation.

When the 9-digit social security number is issued, an earnings record is established on an electronic tape (the Summary Earnings Record) in the headquarters office of the Social Security Administration. The earnings record is identified by the individual's social security number, name, age, sex, and race. Covered earnings reported by employers and by the self-employed are posted to the individual's earnings record. The earnings reports are received on a flow basis, generally via the IRS.

The CWHS System. The CWHS system was initially designed in the late thirties to provide information for program purposes, taking into consideration factors such as random selection, minimum cost, and the Social Security Administration's system for posting earnings.

The CWHS is a stratified systematic cluster sample. A stratified type of design was developed to insure the presence of data for all parts of the country. Clustering of earnings records within strata results from the sample selection procedure—the use of specified combinations of digits in the serial portion of the social security number.

Initially a 20 percent sample, today it is a 1-percent longitudinal sample of people reported with covered earnings.

Population Sampled and Population of Interest. The population sampled, referred to in the tables as 100 percent data, is the set of earnings records to which 1972 earnings (wage and/or self-employment) were posted through September 1973--the processing cutoff date.

Compared with the population of interest, that is, all workers and earnings to be reported for 1972, the population sampled represents about 97.6 percent of all workers and 97.3 percent of all earnings.

Population Sampled, 1972

Race	Total	Men[1/]	Women
Total	93,914,855	57,253,658	36,661,197
White	79,807,402	49,042,867	30,764,535
Negro	9,995,624	5,686,137	4,309,487
Other	2,015,117	1,249,424	765,693
Unknown	2,096,712	1,275,230	821,482

1/ Includes 134,931 with sex unknown in the earnings record.

Sampling Frame. The sampling frame is the updated Summary Earnings Record. This is an electronic tape file of all earnings records; a record is established for each social security number issued. It contained about 225 million records, equivalent to the social security numbers issued.

Sampling Unit. Earnings records clustered by a specific 3-digit combinations in the serial portion of the social security number comprise each sampling unit. The data, however, refer to persons. The distinction arises because of the possibility of people with more than one social security number. The effect of this is believed to be very small, particularly for data for one year.

Sample Selection. Only the serial portion of the social security number is used for sample selection. Consequently, each block of 9,999 social security numbers (serials of 0000 are not issued) with the same area-group combination may be considered a separate stratum.

Within each stratum, the sampling units to comprise the sample are identified by either of two specific digits in the sixth place, and one of five pairs of 100 possible 2-digit combinations in the eighth and ninth positions. For 1972, there are 940,782 earnings records or people in the sample.

Estimation. Estimates of numbers of persons or amount of earnings were calculated by multiplying sample data by 100. Estimates of means are unweighted sample mean.

Actual Sampling Error. It's at this point in a description of a sample design that sampling error is usually covered. Here, however, I'll be specific about the use of the term "actual sampling error."

A difference between a parameter and its sample estimate reflects sampling and any non-sampling errors. As such, it represents the total error. With reference to the population sampled, a review of processing procedures, estimating techniques, and multiple social security numbers indicates that differences arising from factors other than sampling errors for a given year are minimal for most items. (An exception is "Unknown" age.) Consequently, the differences have been termed, "actual sampling error."

Study Results. A comparison of 100 percent and 1 percent data for total number of workers, total taxable earnings, and mean taxable earnings, by sex, indicates that the sample estimates are within 1/3 of 1 percent of the corresponding parameters (Total line of tables 1, 2, 3).

Variation by 5-Year Age Groups. If we examine the detail of these tables, i.e., by 5-year age groupings, the actual sampling error increases, but generally remains below 2 percent, except for age "unknown."

In comparing population and sample data by race as well as sex and 5-year age groups, sample size must be considered. There were 799,115 white workers in the sample and 99,965 Negro workers. Since white workers comprise 85 percent of all workers in the sample, the experience for all workers largely reflects the experience for white workers. Actual sampling errors for white workers (table 4) are generally less than 2 percent. For white male (not shown) actual sampling error was also generally less than 2 percent; white female (not shown) generally less than 3 percent.

For Negro workers, about 11 percent of the sample workers, actual sampling errors by 5-year age groups (table 5) were generally less than 4 percent, as was true for men (not shown) and women (not shown).

Variation by Earnings Interval. Comparisons were also made by $600 earnings intervals, up to the maximum taxable. Actual sampling error was generally less than 1 percent for all workers (table 6) and for other groups not shown: male, female, white men and white women. For Negro workers, actual sampling error was generally under 2 percent, increasing to 3 percent for men and women separately.

Summary. Since some population data are accessible to the Social Security Administration, the actual sampling error (the difference between a sample statistic and its corresponding population parameter) for selected items can be obtained as a byproduct of administrative and statistical operations. Further, the population data can serve as benchmarks.

Population data and 1-percent sample data for 1972 were obtained on a national basis for a limited number of characteristics: Number of workers, taxable earnings, and mean taxable earnings--by sex, race, age, and earnings

interval. The sample results by age and sex (essentially large cells) are generally within 2 percent of corresponding population figures. When race information is incorporated, the sample results are generally within 4 percent.

Footnotes for Tables.

Data reflect taxable earnings for 1972 posted through September 1973. Earnings by interval, 100% data, may not add to total because of rounding. Amount of difference is expressed as percent of 100% data. Male workers, where shown, include unknown sex. 0.00% indicates less than 0.605%.

Table 1. All Workers
Comparison of 100% and 1% Data on Number of Workers, Taxable Earnings and Mean Taxable Earnings, by Age, 1972

Age	Number of Workers			Taxable Earnings			Mean Taxable Earnings			
	100%	1% Infl	%Diff	100%	1% Infl	%Diff	100%	1%	Diff	%Diff
Total	93,914,855	94,078,200	-.17	$470,821,150,310	$471,843,194,600	-.22	$5,013.28	$5,015.44	$-2.16	-.04
LT 20	11,309,172	11,284,000	.22	15,463,613,810	15,493,431,600	-.19	1,367.35	1,373.04	-5.69	-.42
20-24	15,129,265	15,173,300	-.29	56,463,211,810	56,513,290,600	-.09	3,732.05	3,724.52	7.53	.20
25-29	11,674,366	11,731,300	-.49	64,600,846,358	64,935,217,700	-.52	5,533.56	5,535.21	-1.65	-.03
30-34	8,948,197	9,039,100	-1.02	54,969,227,466	55,372,303,700	-.73	6,143.05	6,125.86	17.19	.28
35-39	7,917,424	7,933,500	-.20	49,440,160,830	49,647,218,900	-.42	6,244.48	6,257.92	-13.45	-.22
40-44	8,041,016	8,072,400	-.39	50,676,234,227	50,766,902,200	-.18	6,302.22	6,288.95	13.27	.21
45-49	8,291,328	8,276,700	.18	52,237,574,614	52,248,566,600	-.02	6,300.27	6,312.73	-12.46	-.20
50-54	7,699,968	7,725,500	-.33	48,019,437,431	48,300,460,100	-.59	6,236.32	6,252.08	-15.77	-.25
55-59	6,512,262	6,474,100	.59	39,626,205,377	39,421,995,700	.52	6,084.86	6,089.19	-4.33	-.07
60-64	4,726,759	4,708,200	.39	26,484,367,955	26,400,421,800	.32	5,603.07	5,607.33	-4.26	-.08
60-61	2,171,520	2,142,900	1.32	12,786,459,731	12,602,998,100	1.43	5,888.25	5,881.28	6.97	.12
62-64	2,555,239	2,565,300	-.39	13,697,908,224	13,797,423,700	-.73	5,360.72	5,378.48	-17.77	-.33
65-69	2,190,526	2,208,000	-.80	8,253,177,928	8,285,881,700	-.40	3,767.67	3,752.66	15.01	.40
70-71	451,586	450,800	.17	1,308,200,671	1,291,152,800	1.30	2,896.90	2,864.14	32.77	1.13
72+	941,263	927,400	1.47	2,802,371,667	2,742,758,500	2.13	2,977.25	2,957.47	19.78	.66
UNK	81,723	73,900	9.57	476,520,166	423,592,700	11.11	5,830.92	5,731.97	98.95	1.70

Table 2. All Male
Comparison of 100% and 1% Data on Number of Workers, Taxable Earnings and Mean Taxable Earnings, by Age, 1972

Age	Number of Workers			Taxable Earnings			Mean Taxable Earnings			
	100%	1% Infl	%Diff	100%	1% Infl	%Diff	100%	1%	Diff	%Diff
Total	57,253,658	57,433,200	-.31	$342,955,303,299	$343,990,802,800	-.30	$5,990.10	$5,989.41	$ .70	.01
LT 20	6,492,961	6,500,400	-.11	10,088,065,672	10,119,787,700	-.31	1,553.69	1,556.79	-3.10	-.20
20-24	8,675,173	8,714,500	-.45	36,949,354,584	37,068,832,700	-.32	4,259.21	4,253.70	5.51	.13
25-29	7,301,923	7,347,300	-.62	47,471,435,411	47,703,273,200	-.49	6,501.22	6,492.63	8.60	.13
30-34	5,809,631	5,866,400	-.98	43,040,267,463	43,355,677,400	-.73	7,408.43	7,390.51	17.93	.24
35-39	5,012,021	5,044,600	-.65	38,084,279,195	38,332,163,800	-.65	7,598.59	7,598.65	-.07	-.00
40-44	4,964,278	4,960,100	.08	37,886,303,435	37,824,978,300	.16	7,631.79	7,625.85	5.94	.08
45-49	4,984,135	5,007,400	-.47	37,803,942,991	37,980,135,100	-.47	7,584.86	7,584.80	.05	.00
50-54	4,642,087	4,656,900	-.32	34,441,644,570	34,658,152,800	-.63	7,419.43	7,442.32	-22.89	-.31
55-59	3,993,589	3,968,000	.64	28,472,914,739	28,373,362,900	.35	7,129.66	7,150.55	-20.89	-.29
60-64	2,972,569	2,962,600	.34	19,192,226,921	19,112,747,400	.41	6,456.44	6,451.34	5.10	.08
60-61	1,354,850	1,340,400	1.07	9,215,716,620	9,109,489,300	1.15	6,802.02	6,796.10	5.92	.09
62-64	1,617,719	1,622,200	-.28	9,976,510,301	10,003,258,100	-.27	6,167.02	6,166.48	.55	.01
65-69	1,402,040	1,422,900	-1.49	6,010,346,234	6,070,406,400	-1.00	4,286.86	4,266.22	20.64	.48
70-71	298,253	295,200	1.02	970,343,953	941,491,900	2.97	3,253.43	3,189.34	64.09	1.97
72+	639,931	630,900	1.41	2,134,259,116	2,099,540,100	1.63	3,335.14	3,327.85	7.29	.22
UNK	65,067	56,000	13.93	409,919,018	350,253,100	14.56	6,299.95	6,254.52	45.43	.72

Table 3. All Female
Comparison of 100% and 1% Data on Number of Workers, Taxable Earnings and Mean Taxable Earnings, by Age, 1972

Age	Number of Workers 100%	1% Infl	%Diff	Taxable Earnings 100%	1% Infl	%Diff	Mean Taxable Earnings 100%	1%	Diff	%Diff
Total	36,661,197	36,645,000	.04	$127,865,847,010	$127,852,391,800	.01	$3,487.77	$3,488.95	$-1.17	-.03
LT 20	4,816,211	4,783,600	.68	5,375,548,138	5,373,643,900	.04	1,116.14	1,123.35	-7.21	-.65
20-24	6,454,092	6,458,800	-.07	19,513,857,226	19,444,457,900	.36	3,023.49	3,010.54	12.95	.43
25-29	4,372,443	4,384,000	-.26	17,129,410,947	17,231,944,500	-.60	3,917.58	3,930.64	-13.06	-.33
30-24	3,138,566	3,172,700	-1.09	11,928,960,003	12,016,626,300	-.73	3,800.77	3,787.51	13.26	.35
35-39	2,905,403	2,888,900	.57	11,355,881,635	11,315,055,100	.36	3,908.54	3,916.73	-8.20	-.21
40-44	3,076,738	3,112,300	-1.16	12,789,930,792	12,941,923,900	-1.19	4,156.98	4,158.32	-1.34	-.03
45-49	3,307,193	3,269,300	1.15	14,433,631,623	14,268,431,500	1.14	4,364.31	4,364.37	-.05	-.00
50-54	3,057,881	3,068,600	-.35	13,577,792,861	13,642,307,300	-.48	4,440.26	4,445.78	-5.51	-.12
55-59	2,518,673	2,506,100	.50	11,153,290,638	11,048,632,800	.94	4,428.24	4,408.70	19.54	.44
60-64	1,754,190	1,745,600	.49	7,292,141,033	7,287,674,400	.06	4,156.98	4,174.88	-17.90	-.43
60-61	816,670	802,500	1.74	3,570,743,110	3,493,508,800	2.16	4,372.32	4,353.28	19.04	.44
62-64	937,520	943,100	-.60	3,721,397,923	3,794,165,600	-1.96	3,969.41	4,023.08	-53.67	-1.35
65-69	788,486	785,100	.43	2,242,831,694	2,215,475,300	1.22	2,844.48	2,821.90	22.58	.79
70-71	153,333	155,600	-1.48	337,856,718	349,660,900	-3.49	2,203.42	2,247.18	-43.76	-1.99
72+	301,332	296,500	1.60	668,112,552	643,218,400	3.73	2,217.20	2,169.37	47.83	2.16
UNK	16,656	17,900	-7.47	66,601,149	73,339,600	-10.12	3,998.63	4,097.18	-98.56	-2.46

Table 4. White Workers
Comparison of 100% and 1% Data on Number of Workers, Taxable Earnings and Mean Taxable Earnings, by Age, 1972

Age	Number of Workers 100%	1% Infl	%Diff	Taxable Earnings 100%	1% Infl	%Diff	Mean Taxable Earnings 100%	1%	Diff	%Diff
Total	79,807,402	79,911,500	-.13	$414,662,288,398	$415,313,526,900	-.16	$5,195.79	$5,197.17	$-1.38	-.03
LT 20	9,200,242	9,179,500	.23	13,094,398,405	13,120,532,800	-.20	1,423.27	1,429.33	-6.06	-.43
20-24	12,073,805	12,087,200	-.11	46,180,664,545	46,121,999,200	.13	3,824.86	3,815.77	9.09	.24
25-29	9,686,344	9,730,100	-.45	55,119,319,820	55,342,425,100	-.40	5,690.42	5,687.76	2.66	.05
30-34	7,578,131	7,655,800	-1.02	47,949,427,695	48,339,245,700	-.81	6,327.34	6,314.07	13.27	.21
35-39	6,753,017	6,771,000	-.27	43,314,118,029	43,507,276,900	-.45	6,414.04	6,425.53	-11.49	-.18
40-44	6,999,469	7,026,500	-.39	45,236,226,624	45,306,872,000	-.16	6,462.81	6,448.00	14.81	.23
45-49	7,316,568	7,312,700	.05	47,297,971,387	47,288,025,700	.02	6,464.50	6,466.56	-2.06	-.03
50-54	6,852,393	6,878,500	-.38	43,844,808,755	44,130,114,000	-.65	6,398.47	6,415.66	-17.19	-.27
55-59	5,847,509	5,799,100	.83	36,474,432,347	36,259,864,800	.59	6,237.60	6,252.67	-15.07	-.24
60-64	4,257,659	4,235,000	.53	24,470,972,154	24,319,159,400	.62	5,747.52	5,742.42	5.10	.09
60-61	1,961,425	1,930,500	1.58	11,826,651,182	11,586,763,200	2.03	6,029.62	6,001.95	27.67	.46
62-64	2,296,234	2,304,500	-.36	12,644,320,972	12,732,396,200	-.70	5,506.55	5,525.01	-18.47	-.34
65-69	1,949,239	1,961,900	-.65	7,567,513,755	7,582,118,800	-.19	3,882.29	3,864.68	17.61	.45
70-71	399,829	399,200	.16	1,194,056,276	1,184,364,500	.81	2,986.42	2,966.84	19.57	.66
72+	837,147	825,700	1.37	2,578,150,799	2,530,855,600	1.83	3,079.69	3,065.10	14.58	.47
UNK	56,050	49,300	12.04	340,227,808	280,672,400	17.50	6,070.08	5,693.15	376.92	6.21

Table 5. Negro Workers

Comparison of 100% and 1% Data on Number of Workers, Taxable Earnings and Mean Taxable Earnings, by Age, 1972

Age	Number of Workers			Taxable Earnings			Mean Taxable Earnings			
	100%	1% Infl	%Diff	100%	1% Infl	%Diff	100%	1%	Diff	%Diff
Total	9,995,624	9,996,500	-.01	$39,742,775,869	$39,780,480,800	-.09	$3,976.02	$3,979.44	$-3.42	-.09
LT 20	1,357,932	1,346,100	.87	1,433,886,686	1,428,574,500	.37	1,055.93	1,061.27	-5.34	-.51
20-24	1,771,899	1,780,400	-.48	5,688,104,578	5,742,347,800	-.95	3,210.17	3,225.31	-15.14	-.47
25-29	1,300,377	1,312,900	-.96	5,918,186,778	6,004,171,900	-1.45	4,551.13	4,573.21	-22.08	-.49
30-34	1,047,472	1,061,600	-1.35	5,265,993,930	5,261,642,800	.08	5,027.34	4,956.33	71.00	1.41
35-39	920,082	912,600	.81	4,773,943,884	4,723,193,100	1.06	5,188.61	5,175.53	13.07	.25
40-44	837,942	835,600	.28	4,316,004,560	4,314,204,800	.04	5,150.72	5,163.00	-12.28	-.24
45-49	806,214	793,700	1.55	4,009,251,873	4,002,536,300	.17	4,972.94	5,042.88	-69.95	-1.41
50-54	709,965	706,900	.43	3,422,770,382	3,410,859,300	.35	4,821.04	4,825.09	-4.05	-.08
55-59	554,018	557,400	-.61	2,569,472,589	2,541,489,200	1.09	4,637.89	4,559.54	78.34	1.69
60-64	378,869	378,900	-.01	1,575,251,617	1,594,330,700	-1.21	4,157.77	4,207.79	-50.01	-1.20
60-61	170,699	170,900	-.12	758,928,719	788,249,800	-3.86	4,446.01	4,612.35	-166.34	-3.74
62-64	208,170	208,000	.08	816,322,898	806,080,900	1.25	3,921.42	3,875.39	46.04	1.17
65-69	187,681	190,400	-1.45	501,738,636	502,784,800	-.21	2,673.36	2,640.68	32.68	1.22
70-71	38,403	37,600	2.09	76,911,052	70,465,100	8.38	2,002.74	1,874.07	128.66	6.42
72+	73,366	72,800	.77	138,103,477	132,659,300	3.94	1,882.39	1,822.24	60.15	3.20
UNK	11,404	9,600	15.82	53,155,827	51,221,200	3.64	4,661.16	5,335.54	-674.39	-14.47

Table 6. All Workers

Comparison of 100% and 1% Data on Number of Workers, Taxable Earnings and Mean Taxable Earnings, by Earnings Interval, 1972

Earnings Interval	Number of Workers			Taxable Earnings			Mean Taxable Earnings			
	100%	1% Infl	%Diff	100%	1% Infl	%Diff	100%	1%	Diff	%Diff
Total	93,914,855	94,078,200	-.17	$470,821,150,310	$471,843,194,600	-.22	$5,013.28	$5,015.44	$-2.16	-.04
$1-599	11,157,489	11,169,000	-.10	2,973,680,521	2,973,589,100	.00	266.52	266.24	.28	.11
$600-1,199	7,310,555	7,277,600	.45	6,469,905,942	6,442,770,900	.42	885.01	885.29	-.28	-.03
$1,200-1,799	6,237,223	6,271,400	-.55	9,326,805,242	9,374,507,600	-.51	1,495.35	1,494.80	.54	.04
$1,800-2,399	4,837,692	4,825,800	.25	10,107,169,476	10,079,045,400	.28	2,089.25	2,088.58	.68	.03
$2,400-2,999	4,267,436	4,265,100	.05	11,494,375,645	11,486,190,600	.07	2,693.51	2,693.06	.44	.02
$3,000-3,599	4,164,413	4,177,900	-.32	13,736,391,106	13,776,405,300	-.29	3,298.52	3,297.45	1.07	.03
$3,600-4,199	4,532,254	4,540,300	-.18	17,708,856,823	17,744,521,000	-.20	3,907.30	3,908.23	-.93	-.02
$4,200-4,799	4,328,634	4,360,000	-.72	19,455,959,050	19,599,324,400	-.74	4,494.71	4,495.26	-.55	-.01
$4,800-5,399	4,125,550	4,138,900	-.32	21,024,985,811	21,093,909,000	-.33	5,096.29	5,096.50	-.21	-.00
$5,400-5,999	3,811,753	3,850,800	-1.02	21,710,053,890	21,930,430,000	-1.02	5,695.56	5,695.03	.52	.01
$6,000-6,599	3,695,444	3,723,400	-.76	23,264,318,939	23,437,581,600	-.74	6,295.41	6,294.67	.73	.01
$6,600-7,199	3,377,574	3,379,600	-.06	23,285,910,076	23,308,251,300	-.10	6,894.27	6,896.75	-2.48	-.04
$7,200-7,799	3,169,282	3,159,300	.31	23,742,677,854	23,665,690,700	.32	7,491.50	7,490.80	.70	.01
$7,800-8,399	2,977,271	2,974,900	.08	24,092,886,505	24,077,771,800	.06	8,092.27	8,093.64	-1.37	-.02
$8,400-8,999	2,822,465	2,811,200	.40	24,578,132,020	24,478,103,300	.41	8,708.04	8,707.35	.69	.01
$9,000 PLUS	23,099,820	23,153,000	-.23	217,849,041,410	218,375,102,600	-.24	9,430.77	9,431.83	-1.06	-.01