



Research and Statistics Note

Comparing Earnings Estimates from the 2006 Earnings Public-Use File and the *Annual Statistical Supplement*

By Michael Compson*

Introduction

The Social Security Administration (SSA) recently released the 2006 Earnings Public-Use File (EPUF).¹ The EPUF contains earnings information for individuals drawn from a systematic random 1-percent sample of all Social Security numbers (SSNs) issued before January 2007. EPUF consists of two linkable subfiles. One contains selected demographic and aggregate earnings information for all 4,348,254 individuals in the file, and the second contains annual earnings records for the 3,131,424 individuals who had positive earnings in at least 1 year from 1951 through 2006.²

Evaluating the accuracy of the EPUF estimates was a critical step in developing the data file. Starting with 1939 data, SSA has published annual estimates of the number of workers and the value of the earnings covered under the programs it administers. The estimates first appeared in the *Social Security Yearbook* and, beginning with data for 1949, have been published in the *Annual Statistical Supplement to the Social Security Bulletin* (hereafter referred to simply as the Supplement). The Office of Research, Evaluation, and Statistics (ORES) produces these estimates using the Continuous Work History Sample (CWHs) sampling frame.³

Given that the CWHs and EPUF represent two distinct sampling frames, one expects differences in the earnings estimates derived from each. Besides the different sampling techniques, there are four reasons why the two sets of estimates will differ. First, the two estimates are based on different measures

* Michael Compson is with the Division of Policy Evaluation, Office of Research, Evaluation, and Statistics, Office of Retirement and Disability Policy, Social Security Administration.

Acknowledgments: The author gratefully acknowledges the assistance of many individuals in the process of creating the 2006 Earnings Public-Use File and this note: John Hennessey for graciously sharing his programming and methodological expertise; Russell Hudson for his programming expertise and sharing his vast knowledge of the earnings data; Scott Muller and Greg Diez for sharing programmatic and earnings knowledge; Sirisha Anne, Brenda South, Stu Friedrich, and Randall Miles for their assistance in providing the data extracts used in the process of creating EPUF; Susan Grad, Howard Iams, Hilary Waldron, and Anya Olsen, for their comments on previous drafts of the paper; and finally, Bill Davis and Justin Ronca for their statistical expertise.

The findings and conclusions presented in this paper are those of the author and do not necessarily represent the views of the Social Security Administration.

¹For an introduction to the EPUF, see Compson (2011).

²See appendix for more details on the structure of the two subfiles.

³The CWHs is a longitudinal database produced by ORES for internal research and statistical purposes. SSA is authorized to share the CWHs with the Treasury Department's Offices of Economic Policy and Tax Analysis and with the Congressional Budget Office. For more information about the CWHs, see Buckler (1988) and Smith (1989).

of earnings: The Supplement uses Social Security taxable earnings and EPUF uses capped Social Security taxable earnings. Second, the Supplement estimates are adjusted using factors developed by SSA’s Office of the Chief Actuary (OCACT) to account for delinquent or fraudulent reporting of Form W-2 and Form 1040 Schedule SE information. Third, ORES and OCACT use different methodologies for updating historical estimates. Finally, EPUF removes some individuals and some earnings records (which are set equal to \$0) from the underlying 1-percent sample to “clean” the data and to prevent disclosing personal information.

Selected Abbreviations	
CWHS	Continuous Work History Sample
EPUF	Earnings Public-Use File
ESF	Earnings Suspense File
MEF	Master Earnings File
OCACT	Office of the Chief Actuary
ORES	Office of Research, Evaluation, and Statistics
SSA	Social Security Administration
SSN	Social Security number

This note identifies and explains the differences between data in EPUF and estimates in the Supplement. It first highlights the factors that contribute to expected differences between the two estimates. It then compares EPUF and Supplement estimates, in turn focusing on earnings, number of workers with earnings, median earnings by sex and age group, and the percentage of workers with earnings below the taxable maximum by sex. After accounting for the expected differences, the note finds that remaining differences between EPUF and Supplement estimates are relatively small.

Expected Differences in the Estimates

This discussion distinguishes between the EPUF’s underlying sample and the final EPUF data file. The *underlying sample* refers to a file containing earnings records for 4,413,024 individuals, before data cleaning and disclosure prevention procedures (discussed later) led to the removal of some earnings records. The final EPUF (or, simply, *EPUF*) contains the earnings records for 4,348,254 individuals. The underlying sample and the EPUF use different earnings measures, explained in the following section.

Different Measures of Earnings

All of the earnings data needed to administer the Social Security programs are contained in the Master Earnings File (MEF).⁴ The MEF consists of 20 segments, each containing specific data fields used for various administrative purposes. The Supplement earnings estimates analyzed here are taken from the MEF summary segment using the CWHS sampling frame.⁵ In general, the annual earnings data on the MEF summary segment are a running total of an individual’s earnings up to the taxable maximum for each job in a given year, plus any taxable self-employment income. For the self-employed, “taxable earnings consists of net self-employment income which, when combined with any taxable wages for that individual, is at or below any applicable annual maximum taxable amount” (SSA 2011, G17).

MEF data reflect Social Security taxable earnings; that is, all earnings covered under the program subject to the payroll tax. Note that if an individual has more than one employer in a given year, the amount of earnings in this field may exceed the taxable maximum.⁶

⁴For information on the MEF, see Olsen and Hudson (2009).

⁵Some Supplement tables are based on CWHS annual files. However, this analysis examines Supplement earnings tables based on the MEF 1-percent sample, an extract of earnings data from the MEF summary segment using the CWHS sampling frame, which selects a random sample of records based on certain serial digits of the SSN.

⁶Although other circumstances may account for records with taxable earnings above the taxable maximum, the vast majority of cases involve earnings from multiple employers.

The Supplement and the 1-percent MEF sample that underlies the EPUF use the same earnings measure, Social Security taxable earnings. However, in the final EPUF, earnings data for a given year are capped at the taxable maximum.⁷ Capped Social Security taxable earnings reflect a worker’s covered earnings that are subject to the employee share of the payroll tax.

The following scenario illustrates the differences between the taxable earnings amount in the Supplement tables and the capped taxable earnings contained in EPUF. For a given year, assume the taxable maximum is \$50,000 and an individual has covered earnings from two jobs. If the individual earns \$60,000 in his first job and \$15,000 in a second job, taxable earnings, as shown in the Supplement, would be \$65,000 (\$50,000 from the first job and \$15,000 from the second job). However, the EPUF record would reflect only the capped taxable amount of covered earnings, or the individual’s total covered earnings subject to the employee share of the payroll tax (\$50,000). Given the difference between taxable earnings (Supplement) and capped taxable earnings (EPUF), one would expect the earnings amount in EPUF to be less than the Supplement earnings estimate. The difference between the two measures of earnings is the amount of covered earnings above the taxable maximum earned from multiple jobs, and it accounts for most of the differences between the Supplement and EPUF earnings estimates.

Adjustments to Taxable Earnings

In the Supplement, the estimates of annual taxable earnings and the number of workers with covered earnings reflect adjustments to the raw data pulled from the MEF summary segment. The adjustments attempt to account for two key issues: (1) earnings data for the most recent years are incomplete, and (2) some earnings data reported on W-2 and Schedule SE tax forms may be erroneous or fraudulent.

In general, by the time data are extracted to generate earnings estimates for the Supplement, approximately 98 percent of the current tax year’s earnings data have been posted to the MEF.⁸ To account for the “missing” data, OCACT generates adjustment factors for the number of workers and the amount of taxable earnings in the extract. The adjustment factors are applied to the raw estimates to approximate the final earnings data expected to be posted to the MEF for the current tax year.⁹

In addition, employers may make errors when reporting employees’ Social Security covered earnings, or individuals may use an SSN fraudulently. SSA has a number of procedures that attempt to identify and correct improperly reported earnings information. If these procedures cannot assign earnings information to an SSN, the record is placed in the Earnings Suspense File (ESF). Once the earnings are posted to the ESF, SSA takes additional steps to try to assign the earnings to the appropriate worker. The amount of taxable earnings data posted to the ESF has increased dramatically in recent years (Chart 1), causing a commensurate shortfall in earnings posted to the MEF. OCACT generates adjustment factors to approximate the number of workers and the amount of earnings currently in the ESF that are expected eventually to be posted to the MEF.

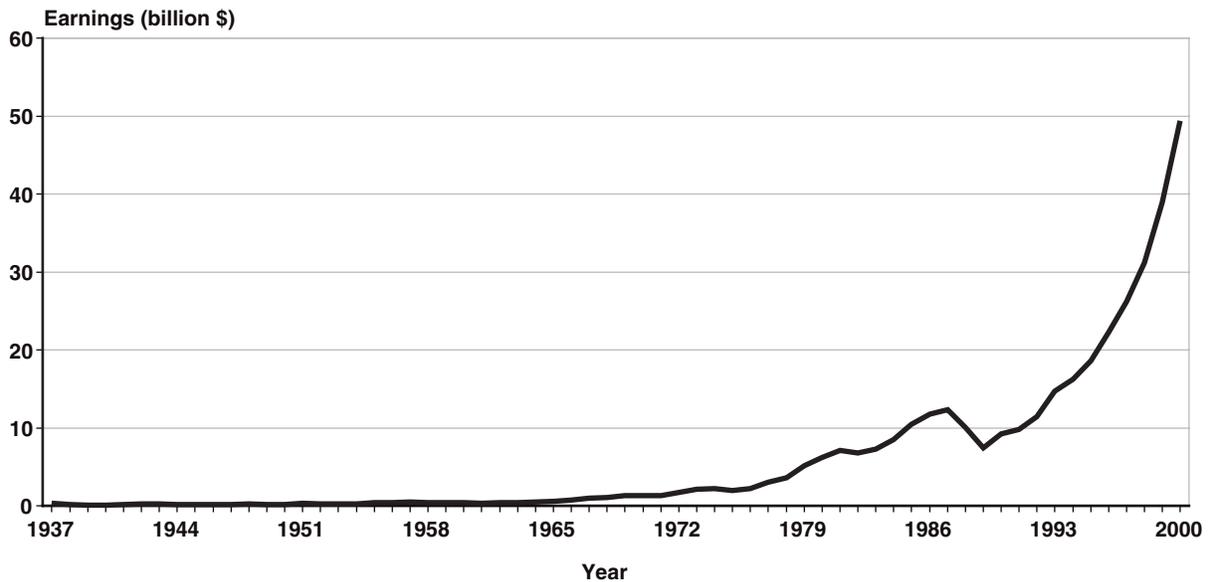
These adjustment factors are used solely to generate earnings estimates in the Supplement, and are not included in the EPUF microdata. Instead, the earnings data underlying the EPUF estimates reflect only the earnings data actually posted on the MEF when the data were extracted.

⁷Capping earnings at the taxable maximum in a given year eliminates the need to top-code this data field.

⁸Posting all the information from the W-2s and selected information from Schedule SE is a massive annual undertaking. The MEF is continuously updated as additional W-2 and Schedule SE information is reported, or previously reported earnings are corrected. For more details, see Olsen and Hudson (2009).

⁹For example, suppose that the total amount of taxable earnings on the MEF was \$98 and the expected total amount of earnings posted to the MEF was \$100. The adjustment factor in this case would be 1.0204082. ORES would multiply the aggregate earnings on the MEF for this tax year by the adjustment factor to generate an estimate of \$100 in the Supplement.

Chart 1.
Value of earnings in the ESF, 1937–2000 (in billions of dollars)



SOURCE: SSA (2002).

The adjustment factors are relatively large for the most recent years' estimates and decrease for each earlier year until the adjustments are minimal. With each passing year, the number of additional earnings data items posted to the MEF for a given tax year decreases. With the growing size of the ESF, one would expect to see greater differences between the Supplement estimates and the EPUF estimates for the most recent years, because both the adjustment factors and the amount of earnings not yet reported on the MEF (thus missing from EPUF) are increasing.

Differences in Historical Estimates

Data published in the Supplement reflect OCACT estimates of worker counts and covered earnings amounts for all earnings years. When generating the current-year estimates of taxable earnings, ORES revises the latest 3 years of estimates and considers OCACT estimates for all prior years to be relatively unchanged. Thus, Supplement estimates for all but the last 3 years are frozen and do not reflect any W-2 and SE information that may have been posted to the MEF in the intervening years. The differences between the historical estimates in the Supplement and those in EPUF (which include updated earnings data) should be minor.

Data Cleaning and Disclosure Prevention Procedures

In creating EPUF, records for some individuals were removed from the underlying sample because of data "cleaning" or because they were included in an existing public-use data file called the New Beneficiary Data System.¹⁰ In addition, annual earnings for individuals with earnings at ages 14 or younger or 86 or older were "zeroed out" (set equal to \$0) to minimize the risk of personal data disclosure.¹¹

¹⁰ For more details, see Compson (2011).

¹¹ For example, if an individual has annual earnings in each year between ages 12 and 62, the EPUF earnings records would reflect \$0 for ages 12, 13, and 14 years old and all of the individual's other earnings records would remain unchanged.

Comparing the Estimates

These comparisons account for two alternative measures of Social Security covered earnings: taxable earnings, as used in the Supplement and the underlying EPUF sample; and earnings capped at the taxable maximum in a given year, as contained in the final EPUF. Directly comparing final EPUF and Supplement earnings estimates would yield somewhat misleading results because (1) each source uses a distinct measure of earnings and (2) the earnings data in EPUF have been adjusted by data cleaning and disclosure prevention procedures.

In lieu of beginning with a direct comparison, we can compare the estimates for taxable Social Security earnings in the Supplement with those in the underlying EPUF sample. Because these two sources use the same earnings measure, we would expect their estimates to differ only because of (1) differing sampling techniques, (2) OCACT's adjustments to account for delinquent, fraudulent, or erroneous reporting of earnings, and (3) ORES' freezing of historical estimates. Because both sources were created using random sampling, one would expect minimal differences between them. If the OCACT adjustment factors are minimal for all but the most recent years, then one would expect the largest differences for those years. If this comparison reveals substantial differences between the estimates, something is clearly wrong.

After comparing Supplement and underlying EPUF sample estimates, the next step is to isolate the effects of the two key differences in the earnings data between the underlying EPUF sample and the final EPUF. First, EPUF records reflect earnings capped at the taxable maximum in a given year. Second, some earnings records were removed from the underlying sample because of data "cleaning," and some annual earnings records were zeroed out to protect against personal data disclosure. Therefore, EPUF's capped taxable earnings amounts will be lower than the taxable earnings estimates in the underlying EPUF sample, and will thus differ even further from the Supplement estimates.

Finally, having established the context of the differences incrementally, we can compare EPUF and Supplement earnings estimates directly. Those comparisons will examine the differences between taxable and capped taxable Social Security earnings and the effects of the data cleaning and disclosure prevention procedures.

Taxable Earnings in the Underlying EPUF Sample

Table 1 compares the taxable earnings in the underlying EPUF sample with the taxable earnings in Table 4.B1 in the 2008 Supplement.¹² Alongside columns respectively presenting estimates from the Supplement and the underlying EPUF sample, a third column expresses the underlying EPUF sample estimate as a percentage of the Supplement estimate. For 1951–1979 and 1988–1999, the underlying EPUF sample estimates equal at least 99 percent of the Supplement estimates. For 4 years between 1980 and 1987, and in each year 2000 through 2004, the underlying EPUF sample estimate drops to between 98 percent and 99 percent of the Supplement estimate. As expected, the most recent years reflect the largest differences between the two.

Chart 2, which graphs the *difference* between the estimates expressed as a percentage of the Supplement estimate, shows relatively small differences between the estimates for most years. Except for 1967 and 1977, the estimates for 1951–1979 differ by less than one-half of one percentage point. For 1980–1988, the differences between the estimates are much more volatile and depart from the 1951–1979 trend line. This observation might be due to the transition from quarterly to annual wage reporting for

¹² Supplement figures cited in this note are primarily from the 2008 edition, the most recent Supplement consulted for this analysis.

Table 1.
Comparing Supplement and underlying EPUF sample estimates: Taxable earnings, 1951–2006

Year	Supplement (millions of dollars)	Underlying EPUF sample ^a (millions of dollars)	Underlying EPUF sample estimate as a percentage of the Supplement estimate
1951	120,770	120,685	99.93
1952	128,640	128,780	100.11
1953	135,870	135,793	99.94
1954	133,520	133,418	99.92
1955	157,540	157,049	99.69
1956	170,720	170,557	99.90
1957	181,380	181,804	100.23
1958	180,720	180,134	99.68
1959	202,310	201,490	99.59
1960	207,000	206,733	99.87
1961	209,640	209,146	99.76
1962	219,050	218,722	99.85
1963	225,550	225,218	99.85
1964	236,390	235,980	99.83
1965	250,730	249,802	99.63
1966	312,540	311,661	99.72
1967	329,960	327,447	99.24
1968	375,840	374,553	99.66
1969	402,550	401,416	99.72
1970	415,600	414,246	99.67
1971	426,960	425,568	99.67
1972	484,110	482,467	99.66
1973	561,850	559,107	99.51
1974	636,760	634,040	99.57
1975	664,660	663,172	99.78
1976	737,700	735,122	99.65
1977	816,550	812,115	99.46
1978	915,600	912,637	99.68
1979	1,067,000	1,066,210	99.93
1980	1,180,700	1,168,010	98.93
1981	1,294,100	1,289,800	99.67
1982	1,365,300	1,353,330	99.12
1983	1,454,100	1,440,070	99.04
1984	1,608,800	1,582,920	98.39
1985	1,722,600	1,705,990	99.04
1986	1,844,400	1,818,070	98.57
1987	1,960,000	1,939,170	98.94
1988	2,088,400	2,082,980	99.74
1989	2,239,500	2,224,970	99.35
1990	2,358,000	2,345,130	99.45
1991	2,422,500	2,407,820	99.39
1992	2,532,900	2,518,450	99.43
1993	2,636,100	2,617,920	99.31
1994	2,785,200	2,770,150	99.46
1995	2,919,100	2,902,080	99.42

(Continued)

Table 1.
Comparing Supplement and underlying EPUF sample estimates: Taxable earnings,
1951–2006—Continued

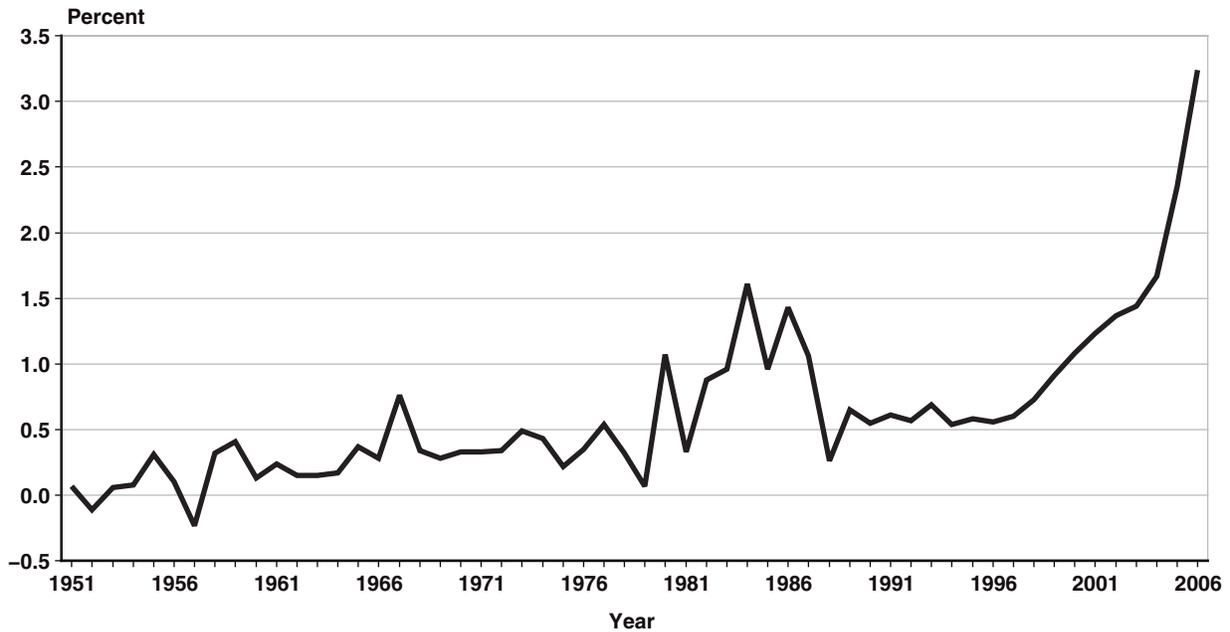
Year	Supplement (millions of dollars)	Underlying EPUF sample ^a (millions of dollars)	Underlying EPUF sample estimate as a percentage of the Supplement estimate
1996	3,073,500	3,056,440	99.44
1997	3,285,000	3,265,320	99.40
1998	3,524,900	3,499,190	99.27
1999	3,749,600	3,715,600	99.09
2000	4,008,500	3,965,170	98.92
2001	4,167,900	4,116,730	98.77
2002	4,250,100	4,191,910	98.63
2003	4,355,000	4,292,200	98.56
2004	^b 4,553,400	4,477,300	98.33
2005	^b 4,765,900	4,653,930	97.65
2006	^b 5,047,755	4,884,310	96.76

SOURCES: SSA (2009, table 4.B1) and author's calculations using underlying EPUF sample.

a. Weighted estimates.

b. 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 2.
Percentage point difference between Supplement and underlying EPUF sample estimates of taxable earnings, 1951–2006



SOURCES: SSA (2009, Table 4.B1) and author's calculations using underlying EPUF sample.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

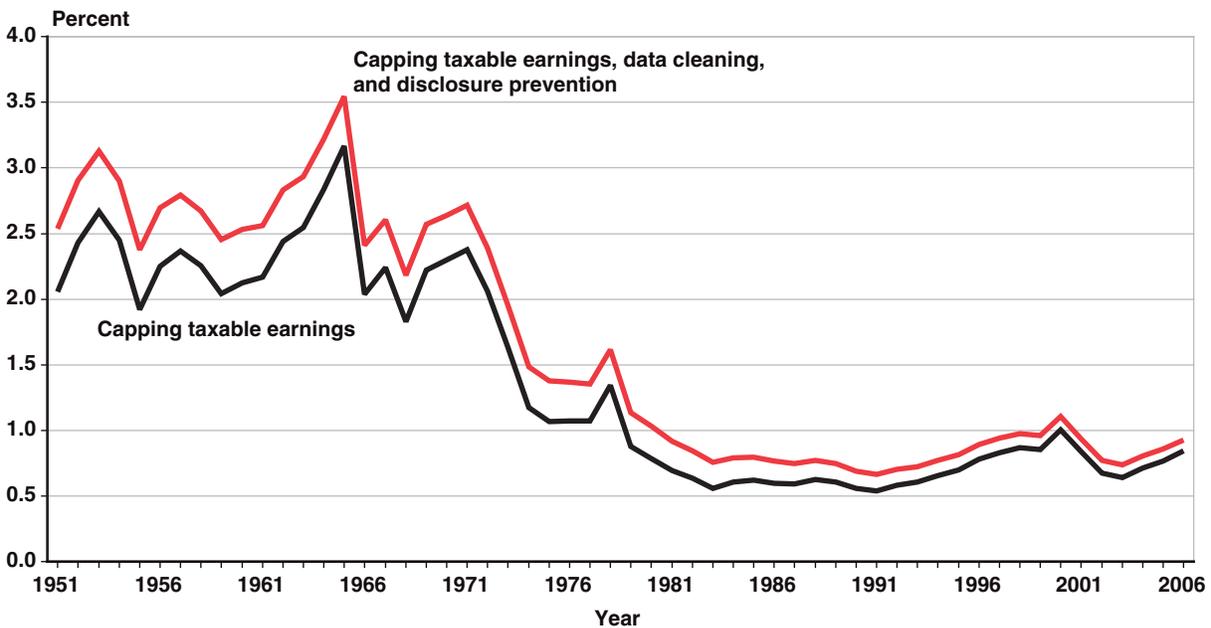
tax year 1978, or to the substantial growth in earnings records assigned to the ESF during that period (Chart 1). It is possible that fewer earnings from the ESF were posted to the MEF during these years than had been expected.¹³ Although the variance in annual estimates from 1980 to 1988 is two to three times that seen in the other years between 1951 and 1997, the differences are still relatively small, only once exceeding 1.5 percentage points.

From 1989 through 1997, the difference in estimates is nearly stationary at one-half of one percentage point. However, from 1998 to 2004, there is a steady increase in the gap between the taxable earnings estimates in the underlying EPUF sample and the Supplement. One possible explanation for the growing gap is the increase in earnings assigned to the ESF that are not recorded in the MEF but are reflected in the Supplement estimates. Also, the difference between the estimates for the two most recent years (2005 and 2006) is much larger because the data in the MEF for those years are incomplete. These findings support the initial expectations about differences between the estimates.

Capped Taxable Earnings in EPUF

Chart 3 presents the percentage of earnings removed from the underlying EPUF sample due to capping earnings at the annual taxable maximum, removing records from the file for data cleaning, and zeroing out some annual earnings values because of data disclosure concerns.¹⁴ The bottom line in Chart 3 reveals that most of the earnings removed from the underlying EPUF sample are the result of capping earnings at the taxable maximum, as opposed to the data cleaning and disclosure prevention procedures (the distance between the lines).

Chart 3.
Percentage of taxable earnings removed from the underlying EPUF sample, by reason, 1951–2006



SOURCE: Author's calculations using the underlying EPUF sample.

¹³ As noted earlier, Supplement estimates make use of OCACT adjustment factors for the number of workers and the amount of earnings reported to the ESF. Those adjustment factors are beyond the scope of this analysis.

¹⁴ Appendix Table A-1 contains the data for Chart 3.

The amount of earnings removed from the underlying EPUF sample expressed as a percentage of total earnings in that sample (top line in Chart 3) starts at 2.5 percent for 1951 and peaks at just over 3.5 percent for 1965. Beginning with 1966, there is a clear downward trend in the percentage of earnings removed from the underlying sample until 1983 (0.8 percent). From 1984 through 2006, the percentage of earnings removed is less than 1 percent, with the single exception of 2000.

Comparing EPUF Capped Taxable Earnings and Supplement Taxable Earnings Estimates

Table 2 shows taxable earnings estimates from Supplement Table 4.B1, weighted capped taxable earnings from the final EPUF, and the latter expressed as a percentage of the former. For years with complete data available, the percentages range from a low of 96.08 percent in 1965 to a high of 98.94 percent in 1988. As expected, the percentages are lower in years with incomplete data, especially 2006. The EPUF estimates are less than 97 percent of the Supplement estimates in only 6 years, with 1971 being the most recent before 2005.

Table 2.
Comparing Supplement taxable earnings estimates with EPUF capped taxable earnings estimates for 1951–2006

Year	Supplement taxable earnings (millions of dollars)	EPUF capped taxable earnings ^a (millions of dollars)	EPUF estimate as a percentage of the Supplement estimate
1951	120,770	117,612	97.39
1952	128,640	125,022	97.19
1953	135,870	131,531	96.81
1954	133,520	129,544	97.02
1955	157,540	153,306	97.31
1956	170,720	165,936	97.20
1957	181,380	176,699	97.42
1958	180,720	175,292	97.00
1959	202,310	196,513	97.13
1960	207,000	201,464	97.33
1961	209,640	203,746	97.19
1962	219,050	212,491	97.01
1963	225,550	218,571	96.91
1964	236,390	228,341	96.60
1965	250,730	240,891	96.08
1966	312,540	304,076	97.29
1967	329,960	318,841	96.63
1968	375,840	366,269	97.45
1969	402,550	390,979	97.13
1970	415,600	403,196	97.02
1971	426,960	413,893	96.94
1972	484,110	470,758	97.24
1973	561,850	547,967	97.53
1974	636,760	624,356	98.05
1975	664,660	653,777	98.36
1976	737,700	724,777	98.25
1977	816,550	800,761	98.07
1978	915,600	897,444	98.02
1979	1,067,000	1,053,555	98.74
1980	1,180,700	1,155,400	97.86

(Continued)

Table 2.
Comparing Supplement taxable earnings estimates with EPUF capped taxable earnings estimates for 1951–2006—Continued

Year	Supplement taxable earnings (millions of dollars)	EPUF capped taxable earnings ^a (millions of dollars)	EPUF estimate as a percentage of the Supplement estimate
1981	1,294,100	1,277,422	98.71
1982	1,365,300	1,341,341	98.25
1983	1,454,100	1,428,558	98.24
1984	1,608,800	1,569,718	97.57
1985	1,722,600	1,691,658	98.20
1986	1,844,400	1,803,448	97.78
1987	1,960,000	1,923,928	98.16
1988	2,088,400	2,066,191	98.94
1989	2,239,500	2,207,592	98.58
1990	2,358,000	2,328,233	98.74
1991	2,422,500	2,391,171	98.71
1992	2,532,900	2,500,096	98.70
1993	2,636,100	2,598,236	98.56
1994	2,785,200	2,748,015	98.66
1995	2,919,100	2,877,705	98.58
1996	3,073,500	3,028,315	98.53
1997	3,285,000	3,233,638	98.44
1998	3,524,900	3,463,966	98.27
1999	3,749,600	3,678,614	98.11
2000	4,008,500	3,920,050	97.79
2001	4,167,900	4,076,775	97.81
2002	4,250,100	4,158,184	97.84
2003	4,355,000	4,259,092	97.80
2004	^b 4,553,400	4,439,555	97.50
2005	^b 4,765,900	4,612,334	96.78
2006	^b 5,047,755	4,837,317	95.83

SOURCES: SSA (2009, table 4.B1) and author's calculations using EPUF.

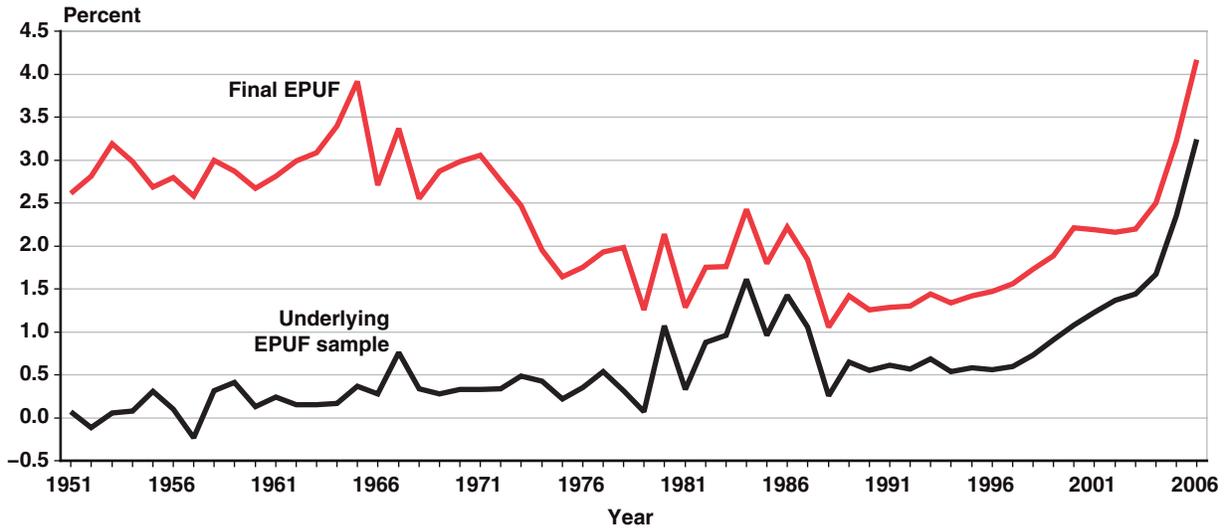
a. Weighted estimates.

b. 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 4 illustrates the effect of changing the measurement from the taxable earnings used in the underlying EPUF sample to the capped taxable earnings used in the final EPUF. Chart 4's top line shows the percentage point difference between EPUF and Supplement earnings estimates and its bottom line shows the percentage point difference between underlying EPUF sample and Supplement earnings estimates (from Chart 2). Chart 4 provides several points of interest. First, the lines differ widely from 1951 to 1980. Second, the volatility in the differences between the two estimates during 1980–1987 occurs for both earnings measurements, as the two lines move in roughly parallel patterns. Third, beginning in 1981, the gap between the two lines narrows and remains consistent thereafter.

In Chart 5, the black line tracks the spread between the lines shown in Chart 4; that is, it shows the percentage-point difference between EPUF's capped taxable earnings and the Supplement's estimate *minus* the percentage-point difference between the underlying EPUF sample's taxable earnings and the Supplement estimate. The differences are relatively small and have narrowed considerably over time. Specifically, the gap peaks at 3.6 percentage points in 1965 and drops to just over 1 percentage point in 1980. From 1981 through 2006, the gap remains steady between 0.7 and 1.1 percentage points.

Chart 4.
Comparing earnings estimates: Percentage point differences between underlying EPUF sample and Supplement estimates and between final EPUF and Supplement estimates: 1951–2006



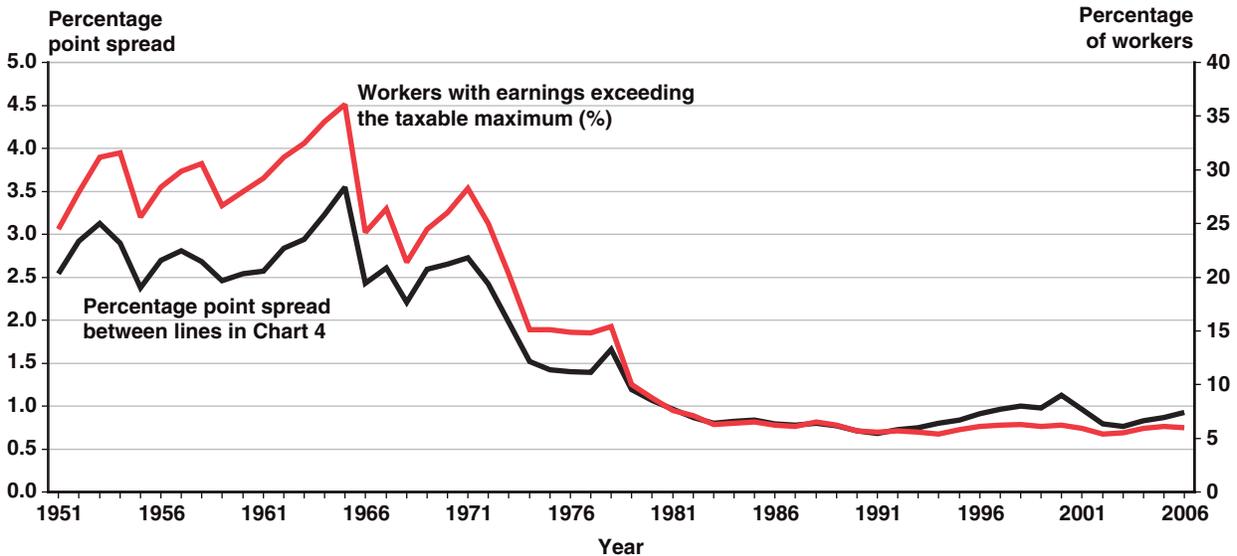
SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, various editions; and author’s calculations using EPUF underlying sample and final EPUF.

NOTES: Supplement and underlying EPUF sample use taxable earnings; final EPUF uses capped taxable earnings.

Charted values are the percentage point differences between the indicated estimates and the Supplement estimate.

2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 5.
Comparing the percentage point spread between the differences in estimates in Chart 4 with the proportion of workers whose earnings exceeded the taxable maximum, 1951–2006



SOURCES: SSA (2009, Table 4.B4) and author’s calculations using EPUF and the underlying EPUF sample.

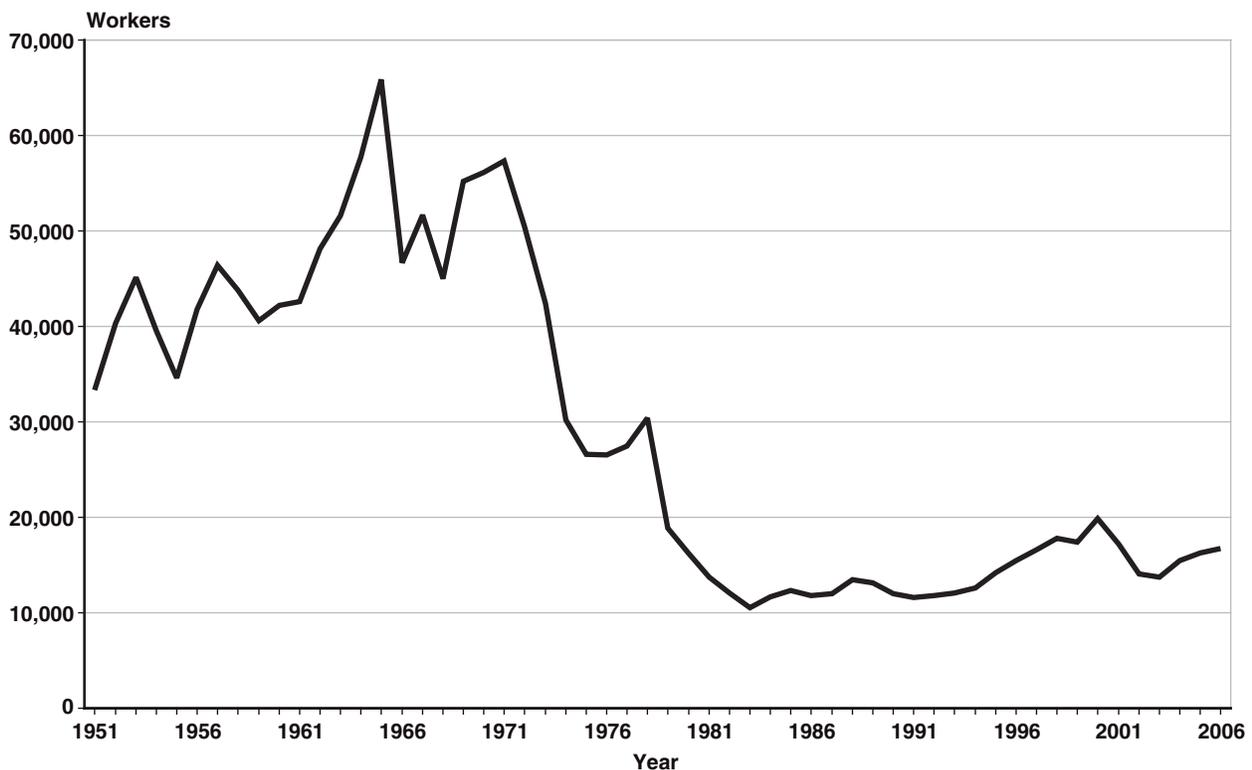
NOTES: “Percentage point spread” equals the percentage-point difference between the final EPUF earnings estimate and the Supplement earnings estimate minus the percentage-point difference between the underlying EPUF sample earnings estimate and the Supplement estimate.

2008 Supplement estimates for 2004–2006 are based on preliminary data.

One possible explanation for the relatively large gap between estimates for 1951–1976 is the much higher percentage of individuals who had earnings above the taxable maximum during those years. As previously noted, the major difference between the taxable earnings in the underlying EPUF sample and the capped earnings in EPUF is that only the latter excludes earnings for workers who have more than one employer and combined earnings above the taxable maximum. As a result, one would expect to see some correlation between the difference in the estimates and the percentage of workers with earnings above the taxable maximum in a given year. The red line in Chart 5 shows the percentage of workers with earnings above the taxable maximum using the scale to the right of the graph. As expected, changes in the percentage of individuals with earnings above the taxable maximum mirror the changes in the differences between taxable earnings and EPUF earnings estimates. The volatility in the percentage of individuals with earnings above the taxable maximum from 1951 to 1970 reflects Congress’ ad hoc adjustments of the taxable maximum during these years. Legislation enacted in 1972 instituted automatic annual increases that took effect with the taxable maximum for 1975.¹⁵

Chart 6 presents the number of workers whose combined taxable earnings from multiple employers exceed the taxable maximum in a given year. The pattern mirrors those of both lines in Chart 5. These findings support initial expectations about capped taxable earnings in EPUF relative to Supplement estimates.

Chart 6.
Number of workers in the underlying EPUF sample with multiple employers and earnings exceeding the taxable maximum, 1951–2006



SOURCE: Author’s calculations using the underlying EPUF sample.

¹⁵ Taxable maximums for 1979 through 1981 were set by legislation. Those for 1990 through 1992 were set using a transitional rule. See SSA (2011, Tables 2.A3 and 2.A18).

Comparing the aggregate earnings estimates derived from the underlying EPUF sample, the final EPUF file, and the Supplement leads to two conclusions: (1) taxable earnings estimates in the underlying EPUF sample and the Supplement do not differ widely; and (2) most of the differences between the final EPUF and Supplement earnings estimates stem from the use of two different measures (taxable and capped taxable earnings) and from OCACT adjustments incorporated in the Supplement estimates to account for delinquent posting of earnings data and potentially fraudulent use of SSNs.

The next sections compare Supplement and final EPUF estimates of the number of workers by sex and age, the median value of taxable earnings by sex and age, and the percentage of workers with earnings below the taxable maximum by sex.

Number of Workers

Supplement Table 4.B3 contains estimates of the number of workers with covered earnings in a given year, by sex. Table 3 compares Supplement and underlying EPUF sample estimates of the number of workers for 1951–2006.¹⁶ Alongside columns presenting the estimates themselves, a third column shows the underlying EPUF sample estimate expressed as a percentage of the Supplement estimate. The estimates differ very little: With one exception (1978), the underlying EPUF sample estimate is within 1 percentage point of the Supplement estimate from 1951 through 1999. As expected, the biggest differences between the estimates occur for the most recent years, when the data are incomplete and OCACT’s adjustment factors play a more prominent role in the Supplement estimates. From 2000 through 2004, the percentages drop below 99 percent; for 2005 and 2006, they drop further, to less than 98 percent.

The next column shows the number of annual earnings records in the underlying EPUF sample removed because of data cleaning or zeroed out to meet data disclosure requirements. The final column reveals that the percentage of underlying EPUF sample records removed or zeroed out is very small, less than 1 percent each year.

Table 4 compares the Supplement and final EPUF estimates of the number of covered workers, with detail by sex. From 1951 through 2004, the final EPUF estimates represent at least 98 percent of the Supplement estimates. As expected, the percentages drop for 2005 and 2006 because of incomplete data and OCACT adjustments.

Workers by Age

Supplement Table 4.B5 shows the estimated number of workers by age group. Unfortunately, some age categories are not defined consistently throughout the 1951–2006 period. Specifically, subcategories for those aged 60 or older from 1951 to 1959 differ from those used from 1960 through 2006. As a result, estimates for those aged 60 or older are shown only for 1960 and later.

Charts 7 and 8 compare EPUF and Supplement estimates of the number of workers with earnings by age group from 1951 to 2006. Both charts show EPUF estimates expressed as a percentage of the Supplement estimate.

¹⁶ Supplement Tables 4.B3, 4.B5, and 4.B6 present annual data only for the most recent years. Data for prior periods are shown only for selected years—specifically, for 1937 and then at 5-year intervals from 1940 until annual coverage begins. Therefore, beginning with Table 3, most of this note’s charts and tables draw data from various editions of the Supplement, always using the most recent edition that presented data for a particular year.

Table 3.**Comparing Supplement and underlying EPUF sample estimates: Number of workers with any earnings during the year, 1951–2006**

Year	Supplement (thousands)	Underlying EPUF sample (thousands)	Underlying EPUF sample estimate as a percentage of the Supplement estimate	Earnings records removed or zeroed out from underlying EPUF sample (thousands)	Underlying EPUF sample earnings records removed or zeroed out (%)
1951	58,120	57,907	99.63	441	0.76
1952	59,580	59,501	99.87	462	0.78
1953	60,840	60,589	99.59	458	0.76
1954	59,610	59,447	99.73	393	0.66
1955	65,200	65,039	99.75	452	0.69
1956	67,610	67,596	99.98	473	0.70
1957	70,590	70,627	100.05	467	0.66
1958	69,770	69,901	100.19	418	0.60
1959	71,700	71,477	99.69	422	0.59
1960	72,530	72,428	99.86	427	0.59
1961	72,820	72,702	99.84	420	0.58
1962	74,280	74,220	99.92	413	0.56
1963	75,540	75,458	99.89	427	0.57
1964	77,430	77,360	99.91	431	0.56
1965	80,680	80,447	99.71	463	0.58
1966	84,600	84,520	99.91	521	0.62
1967	87,040	86,465	99.34	535	0.62
1968	89,380	89,169	99.76	574	0.64
1969	92,060	92,080	100.02	619	0.67
1970	93,090	92,659	99.54	607	0.65
1971	93,340	92,893	99.52	602	0.65
1972	96,240	95,793	99.54	653	0.68
1973	99,830	99,501	99.67	732	0.74
1974	101,330	101,068	99.74	744	0.74
1975	100,200	100,067	99.87	678	0.68
1976	102,600	102,524	99.93	684	0.67
1977	105,800	105,753	99.96	728	0.69
1978	110,600	109,178	98.71	782	0.72
1979	112,700	111,792	99.19	757	0.68
1980	113,000	112,364	99.44	690	0.61
1981	113,000	112,447	99.51	645	0.57
1982	111,800	110,998	99.28	590	0.53
1983	112,100	112,093	99.99	572	0.51
1984	116,300	116,425	100.11	632	0.54
1985	119,800	119,949	100.12	672	0.56
1986	122,900	122,294	99.51	640	0.52
1987	125,600	125,350	99.80	664	0.53
1988	129,600	129,312	99.78	714	0.55
1989	131,700	131,774	100.06	738	0.56
1990	133,600	132,705	99.33	667	0.50
1991	133,000	132,114	99.33	598	0.45
1992	134,000	132,967	99.23	598	0.45
1993	136,100	135,061	99.24	626	0.46
1994	138,200	137,921	99.80	673	0.49
1995	141,000	140,160	99.40	661	0.47

(Continued)

Table 3.
Comparing Supplement and underlying EPUF sample estimates: Number of workers with any earnings during the year, 1951–2006—Continued

Year	Supplement (thousands)	Underlying EPUF sample (thousands)	Underlying EPUF sample estimate as a percentage of the Supplement estimate	Earnings records removed or zeroed out from underlying EPUF sample (thousands)	Underlying EPUF sample earnings records removed or zeroed out (%)
1996	143,400	142,468	99.35	674	0.47
1997	146,145	145,132	99.31	685	0.47
1998	148,786	147,955	99.44	707	0.48
1999	151,333	150,355	99.35	697	0.46
2000	154,732	152,906	98.82	712	0.47
2001	155,416	153,131	98.53	666	0.43
2002	154,893	152,564	98.50	608	0.40
2003	154,576	152,634	98.74	570	0.37
2004	^a 156,250	154,106	98.63	556	0.36
2005	^a 158,913	155,594	97.91	534	0.34
2006	^a 161,205	156,814	97.28	534	0.34

SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B3, various editions; and author's calculations using underlying EPUF sample.

a. 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Table 4.
Comparing Supplement and EPUF estimates: Number of all, male, and female workers with any earnings during the year, 1951–2006

Year	Supplement (thousands)			EPUF ^a (thousands)			EPUF estimate as a percentage of Supplement estimate		
	All ^b	Men	Women	All ^b	Men	Women	All	Men	Women
1951	58,120	38,520	19,600	57,467	38,067	19,366	98.88	98.82	98.80
1952	59,580	39,190	20,390	59,038	38,718	20,284	99.09	98.79	99.48
1953	60,840	39,800	21,040	60,131	39,271	20,825	98.83	98.67	98.98
1954	59,610	39,090	20,520	59,054	38,690	20,332	99.07	98.98	99.08
1955	65,200	43,140	22,060	64,587	42,686	21,862	99.06	98.95	99.10
1956	67,610	44,620	22,990	67,123	44,187	22,893	99.28	99.03	99.58
1957	70,590	47,190	23,400	70,161	46,833	23,286	99.39	99.24	99.51
1958	69,770	46,690	23,080	69,483	46,418	23,029	99.59	99.42	99.78
1959	71,700	47,630	24,070	71,055	47,117	23,904	99.10	98.92	99.31
1960	72,530	47,900	24,630	72,000	47,460	24,509	99.27	99.08	99.51
1961	72,820	47,990	24,830	72,282	47,551	24,701	99.26	99.09	99.48
1962	74,280	48,650	25,360	73,807	48,259	25,519	99.36	99.20	100.63
1963	75,540	49,280	26,260	75,031	48,895	26,108	99.33	99.22	99.42
1964	77,430	50,260	27,170	76,929	49,917	26,983	99.35	99.32	99.31
1965	80,680	51,990	28,690	79,984	51,437	28,518	99.14	98.94	99.40
1966	84,600	53,570	30,870	83,999	53,197	30,774	99.29	99.30	99.69
1967	87,040	54,820	32,220	85,930	54,000	31,901	98.72	98.50	99.01
1968	89,380	55,870	33,510	88,595	55,192	33,373	99.12	98.79	99.59
1969	92,060	56,980	35,080	91,462	56,423	35,007	99.35	99.02	99.79
1970	93,090	57,330	35,760	92,053	56,545	35,475	98.89	98.63	99.20

(Continued)

Table 4.**Comparing Supplement and EPUF estimates: Number of all, male, and female workers with any earnings during the year, 1951–2006—Continued**

Year	Supplement (thousands)			EPUF ^a (thousands)			EPUF estimate as a percentage of Supplement estimate		
	All ^b	Men	Women	All ^b	Men	Women	All	Men	Women
1971	93,340	57,320	36,020	92,291	56,568	35,691	98.88	98.69	99.09
1972	96,240	58,610	37,630	95,141	57,824	37,284	98.86	98.66	99.08
1973	99,830	60,220	39,610	98,769	59,349	39,384	98.94	98.55	99.43
1974	101,330	60,520	40,810	100,324	59,752	40,538	99.01	98.73	99.33
1975	100,200	59,520	40,680	99,389	58,914	40,440	99.19	98.98	99.41
1976	102,600	60,340	42,260	101,839	59,817	41,989	99.26	99.13	99.36
1977	105,800	61,620	44,180	105,025	61,129	43,862	99.27	99.20	99.28
1978	110,600	63,960	46,640	108,397	62,538	45,825	98.01	97.78	98.25
1979	112,700	64,529	48,171	111,035	63,513	47,490	98.52	98.43	98.59
1980	113,000	64,288	48,712	111,674	63,431	48,210	98.83	98.67	98.97
1981	113,000	63,984	49,016	111,802	63,282	48,489	98.94	98.90	98.93
1982	111,800	63,089	48,711	110,408	62,280	48,097	98.75	98.72	98.74
1983	112,100	62,881	49,219	111,520	62,568	48,921	99.48	99.50	99.40
1984	116,300	64,700	51,600	115,793	64,463	51,298	99.56	99.63	99.41
1985	119,800	66,113	53,687	119,277	65,912	53,334	99.56	99.70	99.34
1986	122,900	67,412	55,488	121,654	66,831	54,793	98.99	99.14	98.75
1987	125,600	68,591	57,009	124,686	68,171	56,484	99.27	99.39	99.08
1988	129,600	70,596	59,004	128,598	70,096	58,471	99.23	99.29	99.10
1989	131,700	71,517	60,183	131,036	71,173	59,833	99.50	99.52	99.42
1990	133,600	72,291	61,309	132,038	71,467	60,542	98.83	98.86	98.75
1991	133,000	71,787	61,213	131,516	70,968	60,520	98.88	98.86	98.87
1992	134,000	72,016	61,984	132,369	71,162	61,180	98.78	98.81	98.70
1993	136,100	73,154	62,946	134,435	72,201	62,207	98.78	98.70	98.83
1994	138,200	73,989	64,211	137,247	73,432	63,788	99.31	99.25	99.34
1995	141,000	75,444	65,556	139,500	74,509	64,965	98.94	98.76	99.10
1996	143,400	76,241	67,158	141,794	75,513	66,256	98.88	99.05	98.66
1997	146,145	77,498	68,647	144,448	76,681	67,741	98.84	98.95	98.68
1998	148,786	78,671	70,115	147,247	77,959	69,264	98.97	99.09	98.79
1999	151,333	80,042	71,291	149,657	79,138	70,495	98.89	98.87	98.88
2000	154,732	81,654	73,078	152,194	80,278	71,892	98.36	98.31	98.38
2001	155,416	82,006	73,410	152,465	80,389	72,053	98.10	98.03	98.15
2002	154,893	81,568	73,325	151,956	79,953	71,980	98.10	98.02	98.17
2003	154,576	81,263	73,313	152,064	79,843	72,199	98.37	98.25	98.48
2004	^c 156,250	^c 82,008	^c 74,242	153,551	80,526	73,001	98.27	98.19	98.33
2005	^c 158,913	^c 83,202	^c 75,711	155,060	81,236	73,801	97.58	97.64	97.48
2006	^c 161,205	^c 84,181	^c 77,024	156,280	81,576	74,681	96.94	96.91	96.96

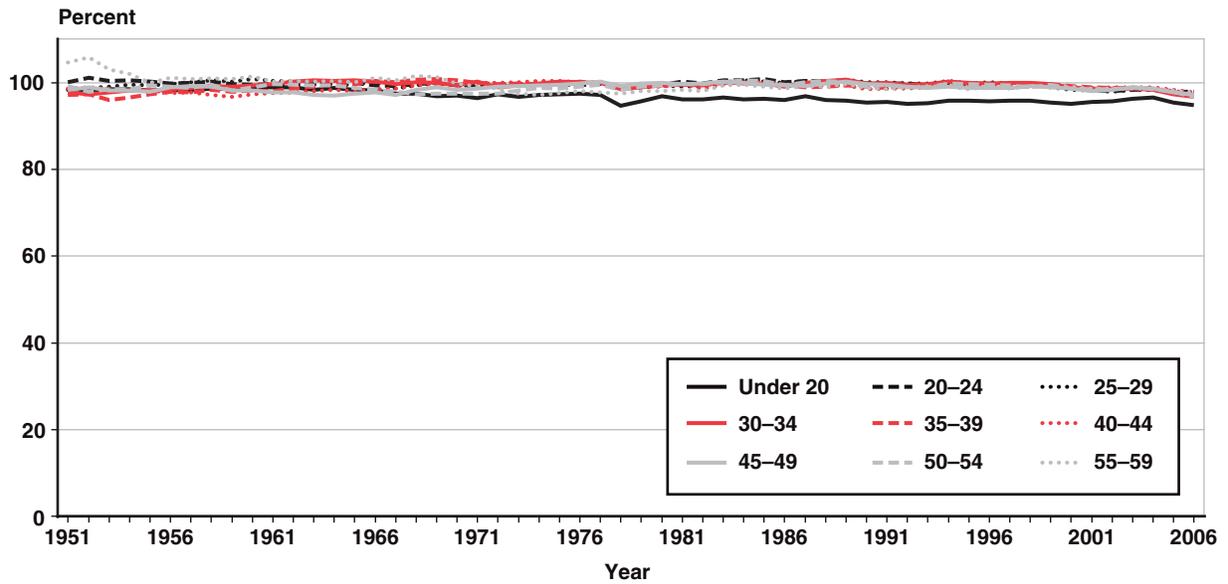
SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B3, various editions; and author's calculations using EPUF.

a. Weighted estimates.

b. Includes a small number of workers whose sex was coded as "unknown."

c. 2008 Supplement estimates for 2004–2006 were based on preliminary data.

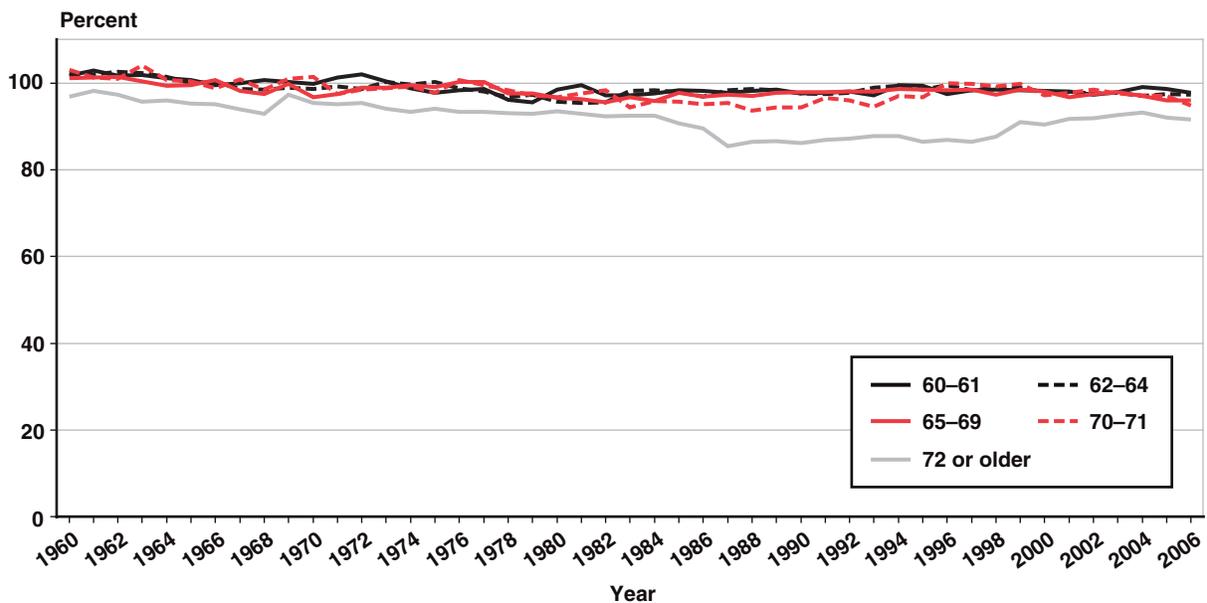
Chart 7.
EPUF estimates of the number of workers as a percentage of the Supplement estimate, workers younger than age 60 by age group, 1951–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B5, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 8.
EPUF estimates of the number of workers as a percentage of the Supplement estimate, workers aged 60 or older by age group, 1960–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B5, various editions; and author's calculations using EPUF.

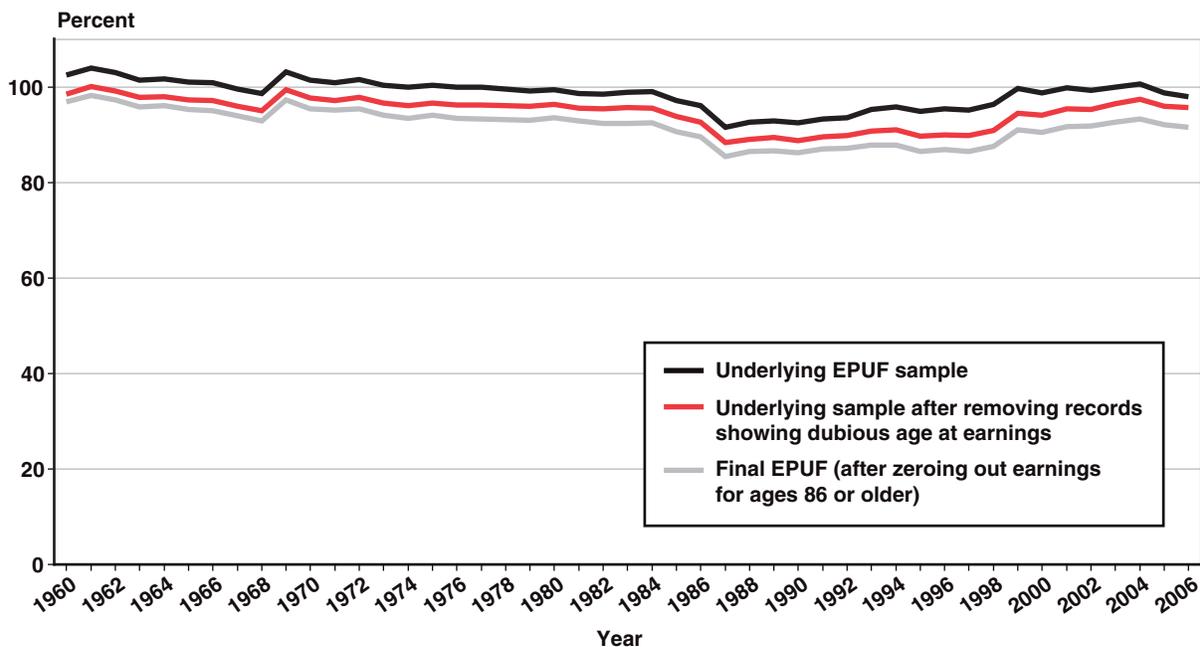
NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 7 looks at workers younger than age 60. In general, the differences between the estimates are very small, although estimates for individuals younger than age 20 clearly diverge starting in 1971. Data disclosure restrictions require EPUF to zero out earnings for those aged 14 or younger, accounting for much of this divergence. When the estimates for the number of workers in this age category are adjusted to include the earnings that were zeroed out, there is virtually no difference between the EPUF and Supplement estimates.

Chart 8 focuses on workers aged 60 or older. Although EPUF and Supplement estimates differ somewhat more for the 60–71 age group than for their younger counterparts, they differ much more for individuals aged 72 or older. For that group, the EPUF estimates are lower than the Supplement estimates across all years, and there is a distinct gap between the estimates for 1985 through 1998. The gap begins to narrow in 1999, but it remains somewhat larger than that for workers aged 60–71. The finding raises two critical questions: Why does this large gap occur only for 1985–1998, and why only for workers aged 72 or older?

The first attempt to answer these questions involves evaluating how the EPUF data cleaning and disclosure prevention requirements affect the estimated number of workers aged 72 or older. Chart 9 presents three measures of workers aged 72 or older as percentages of the Supplement estimate. The top line shows the full underlying EPUF sample estimate. The middle line represents the underlying sample after removing records because of data cleaning (for example, records with dubious age-at-earnings values) and for disclosure prevention (individuals that overlapped with the New Beneficiary Data System). The bottom line, showing the final EPUF after zeroing out earnings for workers aged 86 or older, replicates Chart 8’s line for workers aged 72 or older. The short distance between the bottom line and the middle

Chart 9.
Effects of EPUF data cleaning and disclosure prevention measures: Estimated number of workers aged 72 or older as a percentage of the Supplement estimate, 1960–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B5, various editions; and author’s calculations using EPUF.

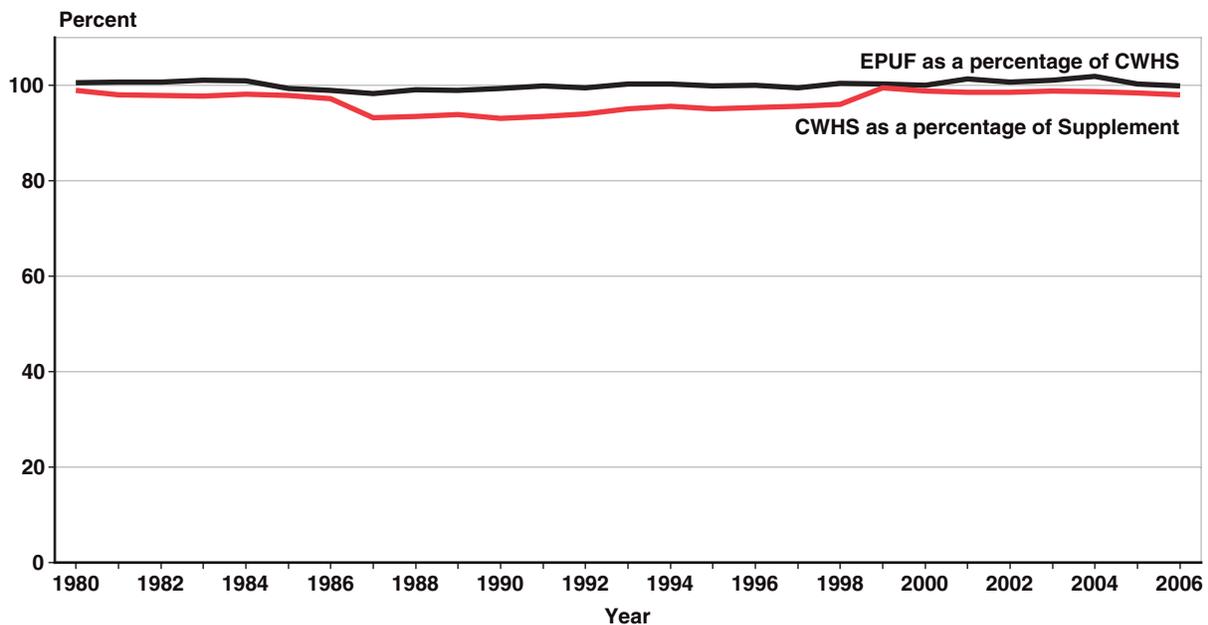
NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

line shows the minimal effect of zeroing out the earnings of individuals aged 86 or older. The distance between the middle line and the top line shows that the effect of removing individuals for data cleaning and disclosure prevention is also generally small, although it is somewhat larger than the effect of zeroing out earnings for individuals aged 86 or older.¹⁷

More significantly, the EPUF estimates (bottom line) and those from the underlying EPUF sample (top line) differ very consistently across the years. Thus, the distinct gap between the EPUF and Supplement estimates from 1985 through 1998 clearly does not result from the data cleaning or the disclosure prevention procedures applied to the underlying EPUF sample. It seems very peculiar that the estimates from the underlying EPUF sample are extremely close to the Supplement estimates except for these particular years. What, then, explains this anomalous gap?

A second approach is to compare the number of workers aged 72 or older in the EPUF with the number of workers in the active file within the 2008 version of the CWHS. The active file contains individuals in the 1-percent CWHS who have had any covered earnings since the program's inception. One would expect these two distinct 1-percent samples to produce very similar estimates of the number of workers aged 72 or older. The top line in Chart 10 shows the number of workers in EPUF expressed as a percentage of the number of workers in the active CWHS file and confirms that the estimates are indeed very similar. The bottom line shows the number of workers in the active CWHS file aged 72 or older as a percentage of the Supplement estimates. The gap between these two ratios is nearly identical to the differences between the EPUF and Supplement estimates for this age group.

Chart 10.
Comparison of estimates of number of workers aged 72 or older: EPUF, CWHS, and Supplement, 1980–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, various editions; and author's calculations using 2008 CWHS active file and EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

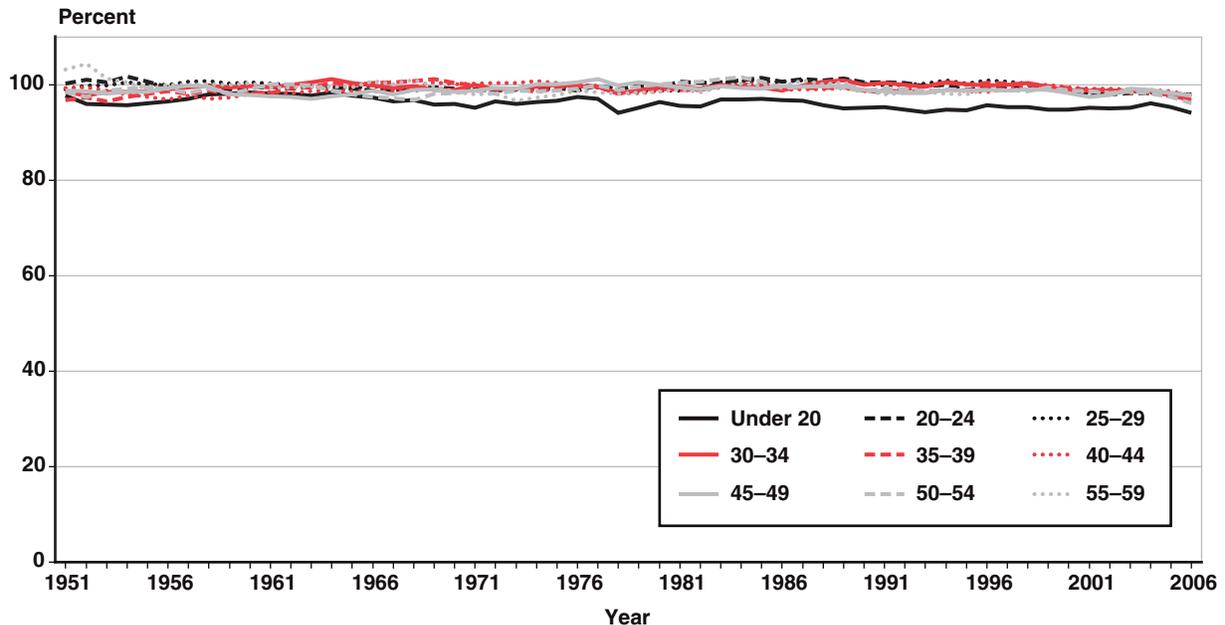
¹⁷ There is no overlap between the individuals removed from the file due to dubious age values and those whose earnings at ages 86 or older were zeroed out. See appendix for more information.

The fact that the estimates from two distinct 1-percent samples are nearly the same indicates that there may be problems with the Supplement estimates for older workers during this period. One possible explanation is that a programming or coding error affected only those workers, and the error was corrected as part of the Y2K adjustments made to the MEF. Nonetheless, the Supplement estimates reflect the earnings data in the MEF at that time. Given the close relationship between EPUF's underlying 1-percent sample and the active CWHS file, the number of workers in EPUF aged 72 or older is presumably correct.

Charts 11–14 compare EPUF and Supplement estimates of the number of workers by age group and sex. For men younger than 60, Chart 11 reveals very little difference between the estimates. As expected, EPUF estimates as a percentage of Supplement estimates decline slightly for the most recent years. As was seen with all workers, estimates of the number of men aged 60 or older (Chart 12) differ more widely than estimates of the number of younger men. The distinct gap between EPUF and Supplement estimates of all workers aged 72 or older from 1985 through 1998 also occurs for men.

For female workers younger than age 60, Chart 13 reveals that EPUF and Supplement estimates differ slightly more than do those of their male counterparts. After 1980, the estimates differ only minimally. As was true for men, EPUF and Supplement estimates for female workers aged 60 or older (Chart 14) differ more than those for younger women. In general, the EPUF estimates of older female workers appear to slightly exceed Supplement estimates from 1960 through 1974 but are lower from 1975 onward.

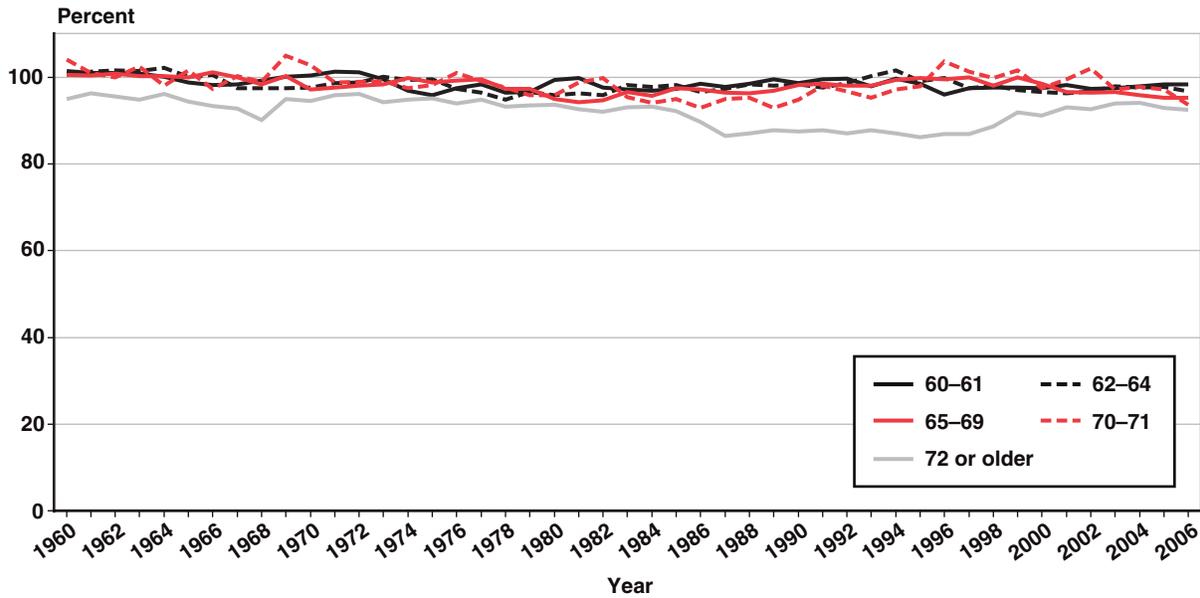
Chart 11.
EPUF estimates of the number of male workers as a percentage of the Supplement estimate, workers younger than 60 by age group, 1951–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B5, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

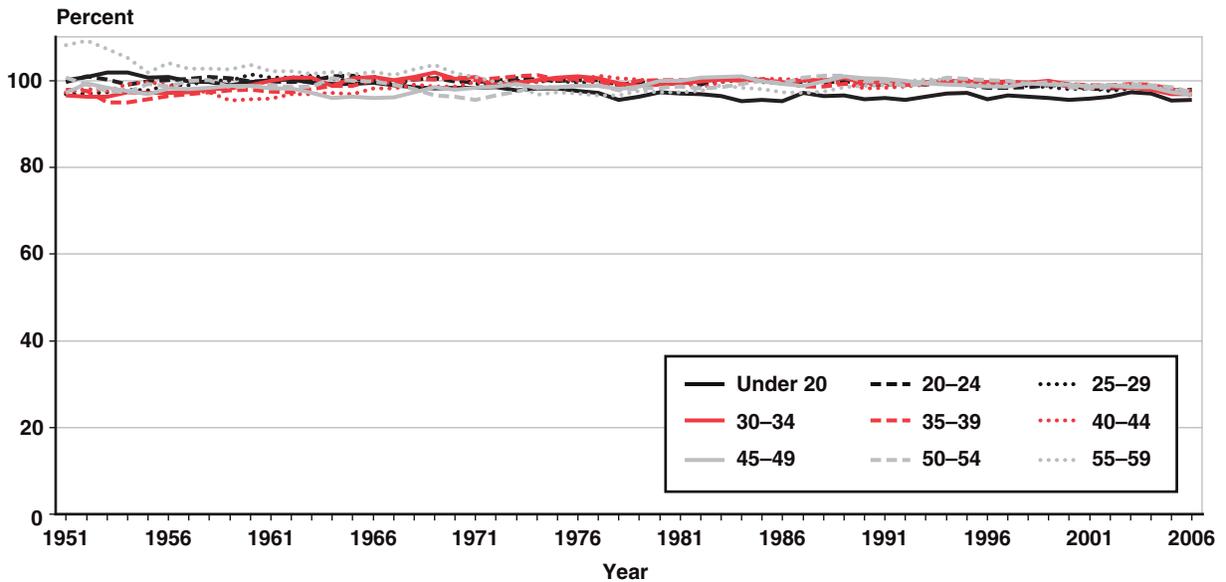
Chart 12.
EPUF estimates of the number of male workers as a percentage of the Supplement estimate, workers aged 60 or older by age group, 1960–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B5, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 13.
EPUF estimates of the number of female workers as a percentage of the Supplement estimate, workers younger than age 60 by age group, 1951–2006

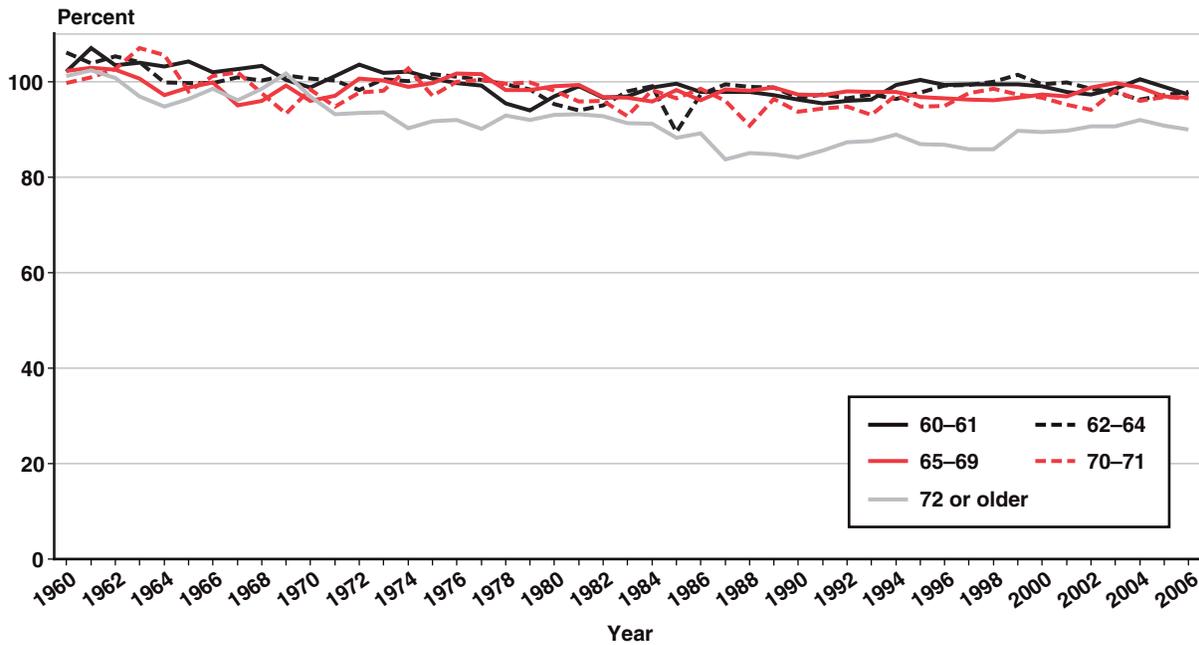


SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B5, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 14.

EPUF estimates of the number of female workers as a percentage of the Supplement estimate, workers aged 60 or older by age group, 1960–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B5, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Median Earnings By Sex and Age

This section compares EPUF estimates of median earnings with those presented in Supplement Table 4.B6. In turns, the discussion examines median earnings for all workers, workers by sex, all workers by age, and then workers by sex and age.

Table 5 compares the median earnings for all, male, and female workers for 1951–2006. The EPUF estimate for all workers is at least 98.44 percent of the Supplement estimate in all years, and in fact slightly exceeds the Supplement estimate for most years. The pattern for men is very similar. For women, the EPUF estimate is much lower than the Supplement estimate in many years; for nine in particular, the EPUF estimate represents less than 98 percent of the Supplement estimate.

Chart 15 reveals the minimal differences between the EPUF and Supplement estimates of median earnings for all workers younger than age 60. Chart 16 shows much more variation for workers aged 60 or older. The greatest variation is for individuals aged 72 or older and it occurs during the same years that showed the greatest differences in the estimated numbers of individuals with earnings (Chart 8).

For men younger than age 60, EPUF median earnings estimates differ substantially from Supplement estimates for 1960–1973 (Chart 17). The greatest difference appears for 1965, when the EPUF estimate is approximately 74 percent of the Supplement estimate for those aged 35–49. How and why do such large differences occur, and why only from 1960 through 1973?¹⁸

¹⁸ As previously noted, the age subcategories used in the Supplement for individuals aged 60 or older are not consistent throughout the 1951–2006 period. For that reason, the analysis is limited to 1960 through 2006.

Table 5.
Comparing Supplement and EPUF estimates: Median earnings for all, male, and female workers, 1951–2006

Year	Supplement (dollars)			EPUF (dollars)			EPUF estimate as a percentage of Supplement estimate		
	All	Men	Women	All	Men	Women	All	Men	Women
1951	2,097	2,838	1,192	2,100	2,900	1,200	100.14	102.18	100.67
1952	2,258	3,046	1,278	2,300	3,100	1,300	101.86	101.77	101.72
1953	2,400	3,275	1,357	2,400	3,300	1,300	100.00	100.76	95.80
1954	2,425	3,263	1,374	2,400	3,300	1,400	98.97	101.13	101.89
1955	2,438	3,315	1,351	2,400	3,300	1,300	98.44	99.55	96.23
1956	2,599	3,546	1,454	2,600	3,500	1,500	100.04	98.70	103.16
1957	2,651	3,538	1,544	2,700	3,600	1,500	101.85	101.75	97.15
1958	2,674	3,516	1,589	2,700	3,500	1,600	100.97	99.54	100.69
1959	2,837	3,783	1,634	2,800	3,800	1,600	98.70	100.45	97.92
1960	2,894	3,879	1,679	2,900	3,900	1,700	100.21	100.54	101.25
1961	2,938	3,936	1,742	2,900	3,900	1,700	98.71	99.09	97.59
1962	3,058	4,132	1,808	3,100	4,100	1,800	101.37	99.23	99.56
1963	3,149	4,266	1,856	3,100	4,300	1,800	98.44	100.80	96.98
1964	3,298	4,480	1,945	3,300	4,500	1,900	100.06	100.45	97.69
1965	3,414	4,685	1,984	3,400	4,754	2,000	99.59	101.47	100.81
1966	3,566	5,010	2,082	3,600	5,000	2,000	100.95	99.80	96.06
1967	3,716	5,208	2,259	3,700	5,200	2,200	99.57	99.85	97.39
1968	3,945	5,546	2,435	4,000	5,600	2,400	101.39	100.97	98.56
1969	4,173	5,933	2,585	4,200	6,000	2,600	100.65	101.13	100.58
1970	4,375	6,180	2,735	4,400	6,200	2,700	100.57	100.32	98.72
1971	4,605	6,475	2,882	4,600	6,500	2,900	99.89	100.39	100.62
1972	4,870	6,923	2,983	4,900	7,000	3,000	100.62	101.11	100.57
1973	5,184	7,473	3,148	5,200	7,500	3,100	100.31	100.36	98.48
1974	5,531	7,972	3,435	5,500	8,000	3,400	99.44	100.35	98.98
1975	5,803	8,250	3,730	5,800	8,300	3,700	99.95	100.61	99.20
1976	6,235	8,883	4,063	6,300	8,900	4,100	101.04	100.19	100.91
1977	6,630	9,489	4,358	6,700	9,600	4,300	101.06	101.17	98.67
1978	7,204	10,279	4,856	7,300	10,400	4,900	101.33	101.18	100.91
1979	7,930	11,258	5,433	7,900	11,300	5,400	99.62	100.37	99.39
1980	8,549	11,963	6,012	8,600	12,000	6,000	100.60	100.31	99.80
1981	9,361	12,941	6,690	9,400	13,000	6,700	100.42	100.46	100.15
1982	9,924	13,318	7,232	9,900	13,300	7,200	99.76	99.86	99.56
1983	10,322	13,687	7,618	10,300	13,700	7,600	99.79	100.09	99.76
1984	10,757	14,360	7,878	10,900	14,500	7,900	101.33	100.97	100.28
1985	11,265	14,959	8,293	11,400	15,100	8,300	101.20	100.94	100.08
1986	11,831	15,579	8,796	11,900	15,600	8,800	100.58	100.13	100.05
1987	12,327	16,073	9,261	12,300	16,100	9,300	99.78	100.17	100.42
1988	12,825	16,613	9,753	12,900	16,600	9,800	100.58	99.92	100.48
1989	13,314	17,014	10,265	13,400	17,200	10,300	100.65	101.09	100.34
1990	13,898	17,582	10,837	14,000	17,800	10,900	100.73	101.24	100.58
1991	14,278	17,765	11,369	14,400	17,900	11,400	100.85	100.76	100.27
1992	14,739	18,208	11,842	14,900	18,400	11,900	101.09	101.05	100.49
1993	15,000	18,430	12,093	15,100	18,700	12,100	100.67	101.47	100.06
1994	15,560	19,249	12,422	15,600	19,400	12,400	100.26	100.78	99.82
1995	16,108	19,907	12,897	16,200	20,000	12,900	100.57	100.47	100.02

(Continued)

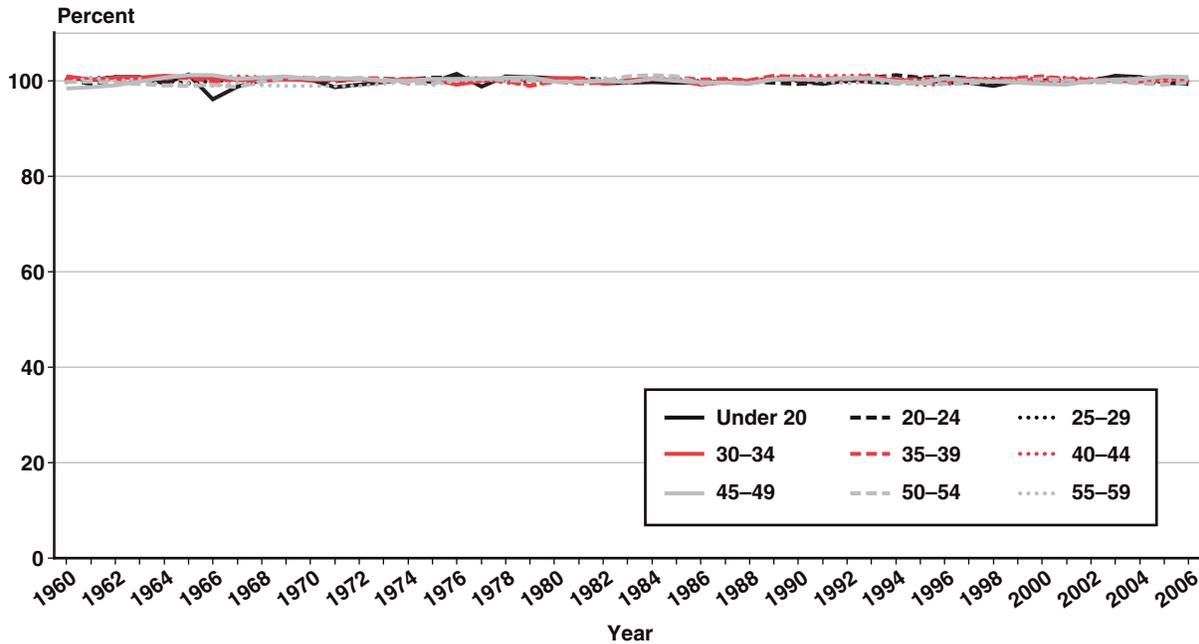
Table 5.
Comparing Supplement and EPUF estimates: Median earnings for all, male, and female workers, 1951–2006—Continued

Year	Supplement (dollars)			EPUF (dollars)			EPUF estimate as a percentage of Supplement estimate		
	All	Men	Women	All	Men	Women	All	Men	Women
1996	16,712	20,779	13,335	16,800	20,800	13,400	100.53	100.10	100.49
1997	17,562	21,814	14,043	17,600	21,900	14,100	100.22	100.39	100.41
1998	18,513	23,028	14,834	18,600	23,100	14,900	100.47	100.31	100.44
1999	19,265	23,927	15,465	19,400	24,100	15,600	100.70	100.72	100.87
2000	20,225	25,032	16,287	20,300	25,200	16,400	100.37	100.67	100.69
2001	20,905	25,643	17,037	21,000	25,700	17,100	100.45	100.22	100.37
2002	21,193	25,765	17,461	21,300	25,900	17,500	100.50	100.52	100.22
2003	21,610	26,173	17,845	21,700	26,300	18,000	100.42	100.49	100.87
2004	^a 22,342	^a 27,074	^a 18,427	22,500	27,200	18,500	100.71	100.47	100.40
2005	^a 22,983	^a 27,895	^a 18,892	23,100	28,000	19,000	100.51	100.38	100.57
2006	^a 23,832	^a 28,916	^a 19,586	24,000	29,100	19,700	100.70	100.64	100.58

SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B6, various editions; and author's calculations using EPUF.

a. 2008 Supplement estimates were based on preliminary data.

Chart 15.
EPUF estimates of median earnings as a percentage of the Supplement estimate, all workers younger than age 60 by age group, 1960–2006

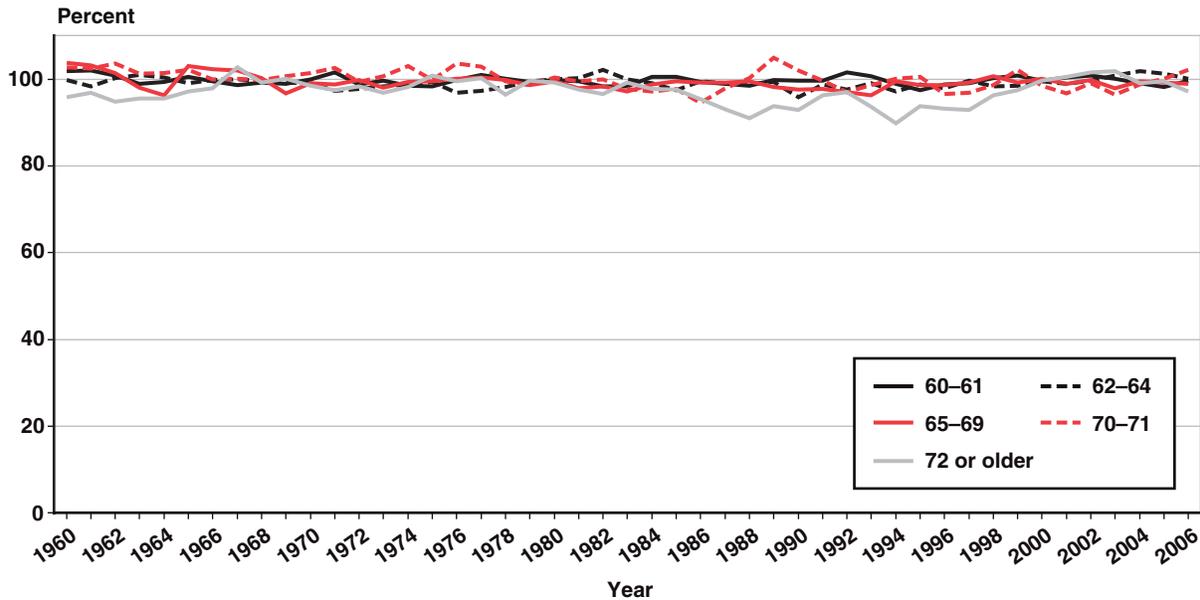


SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B6, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 16.

EPUF estimates of median earnings as a percentage of the Supplement estimate, all workers aged 60 or older by age group, 1960–2006

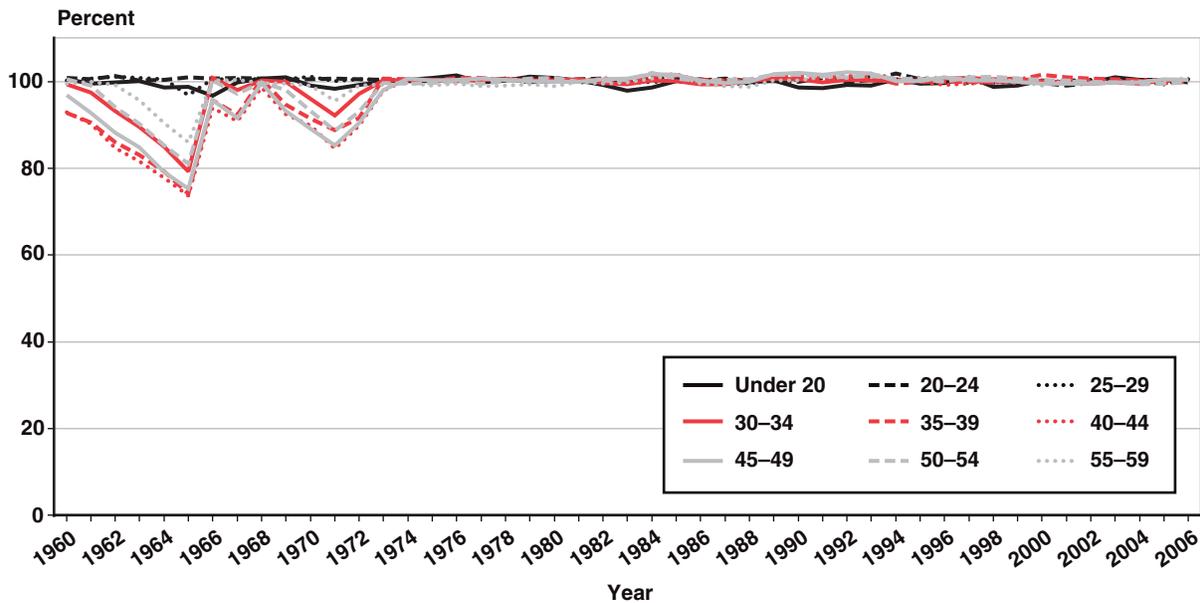


SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B6, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 17.

EPUF estimates of median earnings as a percentage of the Supplement estimate, male workers younger than age 60 by age group, 1960–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B6, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Supplement Table 4.B6 alerts readers that “the amount of median earnings includes estimates above the taxable maximum.” For 1951–1977, those estimates are based on data from earnings reports filed quarterly by employers. SSA began collecting “detailed” earnings information annually from W-2 forms in 1978, but earnings above the taxable maximum were still open to some conjecture through 1993.¹⁹ Starting in 1994, all covered earnings were subject to the Medicare payroll tax; thus, records reflected actual earnings, and estimates of amounts above the taxable maximum were no longer necessary.

If adjustments for earnings above the taxable maximum were made for all years from 1951 through 2006, why do the relatively large differences occur only from 1960 through 1973? Adjustments would affect median earnings only if the preadjustment median value exceeds the taxable maximum. In other words, if the preadjustment median value is less than the taxable maximum, then the additional earnings would not affect the median.

The methodology SSA used through 1994 to estimate earnings above the taxable maximum is not readily available. However, we can determine whether EPUF estimates of median earnings appear reasonable for the years in which they are much lower than Supplement estimates. If EPUF’s median earnings for men in certain age groups is greater than or equal to the taxable maximum in a given year, then we know that any earnings above the taxable maximum will affect the median. Table 6 presents EPUF estimates of median earnings for men as a percentage of the taxable maximum for 1960–1980.²⁰ Shaded

Table 6.
EPUF estimates of median earnings for men as a percentage of the taxable maximum amount, by age group, 1960–1980

Year	19 or younger	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–61	62–64	65–69	70–71	72 or older
1960	12.89	44.42	82.35	100.00	100.00	100.00	100.00	98.44	91.50	83.75	79.85	48.44	24.96	28.44
1961	12.90	43.53	82.98	100.00	100.00	100.00	100.00	100.00	94.64	86.56	81.68	48.57	25.00	29.17
1962	13.43	44.73	88.10	100.00	100.00	100.00	100.00	100.00	100.00	92.19	85.53	49.23	25.00	29.86
1963	12.76	46.19	91.73	100.00	100.00	100.00	100.00	100.00	100.00	95.31	88.95	47.57	24.90	29.00
1964	13.26	50.04	97.70	100.00	100.00	100.00	100.00	100.00	100.00	100.00	94.11	49.84	25.00	29.38
1965	14.61	54.89	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	98.94	54.73	25.00	30.67
1966	12.03	40.86	83.00	98.10	100.00	100.00	100.00	98.25	91.12	83.17	76.10	36.83	22.25	22.92
1967	12.06	39.27	86.99	100.00	100.00	100.00	100.00	100.00	95.02	87.53	80.01	38.30	22.59	24.85
1968	10.94	35.98	79.27	93.79	99.95	100.00	100.00	94.55	86.13	80.19	72.24	32.19	20.55	21.74
1969	11.78	39.51	85.73	100.00	100.00	100.00	100.00	100.00	93.45	85.45	78.02	32.39	20.96	22.83
1970	11.82	42.31	88.49	100.00	100.00	100.00	100.00	100.00	97.55	91.07	81.52	36.14	21.31	23.53
1971	11.84	44.75	90.82	100.00	100.00	100.00	100.00	100.00	100.00	97.12	84.95	37.09	21.53	24.24
1972	11.37	45.91	82.72	100.00	100.00	100.00	100.00	100.00	98.81	90.00	78.99	32.14	18.55	22.20
1973	10.81	42.19	73.24	93.84	100.00	100.00	100.00	99.13	91.04	83.00	70.85	25.66	18.22	20.01
1974	9.70	36.94	62.86	81.75	87.96	90.55	90.83	87.32	80.61	73.20	62.89	20.45	16.65	17.56
1975	8.91	34.58	60.13	79.54	86.27	89.06	89.56	86.78	79.28	71.85	61.60	19.73	16.15	17.09
1976	9.02	34.22	59.48	78.65	87.14	89.26	90.57	87.71	81.08	72.67	60.93	18.61	16.34	16.47
1977	8.84	34.19	58.07	77.66	87.56	89.62	90.79	87.64	81.34	75.14	60.82	18.44	16.03	16.64
1978	9.30	35.17	58.91	77.69	89.00	91.47	92.57	89.86	83.72	76.80	63.09	21.92	16.95	16.77
1979	8.05	29.64	49.30	64.39	74.42	77.11	77.20	75.92	70.85	63.77	53.65	19.03	14.51	14.26
1980	7.22	27.09	45.98	59.99	69.66	72.70	72.88	71.57	67.16	60.41	51.21	18.77	14.09	13.43

SOURCE: Author’s calculations using EPUF.

¹⁹ The cap for covered earnings subject to the Medicare payroll tax was higher than the taxable maximum for the Social Security program in 1992 and 1993, and was removed altogether in 1994.

²⁰ The period 1960–1980 contains the only years in which men’s median taxable earnings in EPUF are equal to the taxable maximum.

cells indicate the years in which one should expect the Supplement medians to be greater than the taxable maximum. Table 7 presents the median earnings values from Supplement Table 4.B6 expressed as a percentage of the taxable maximum; the same cells are shaded as those in Table 6, with one exception (1973, for individuals aged 35–39).

Tables 6 and 7 explain the findings in Chart 17. The EPUF estimates of median earnings for some age categories of male workers are much lower than the Supplement estimates because the EPUF estimates do not account for earnings above the taxable maximum, while the Supplement estimates do. The Supplement’s adjustments for earnings above the taxable maximum increase the estimated median earnings for some age groups in years when the median value exceeds the taxable maximum. However, the median values fall below the taxable maximums for all age groups after 1973, and adjustments for earnings above the taxable maximum no longer affect median values. Thus, EPUF and Supplement median earnings estimates differ minimally from 1973 through 2006.

Chart 18 presents EPUF estimates of median earnings as a percentage of Supplement estimates for male workers aged 60 or older and reveals more divergence than was seen for younger workers, especially for men aged 72 or older. Most of the variation occurs in the same years for which EPUF and Supplement estimates of number of workers vary.

Chart 19 shows minimal differences between EPUF and Supplement median earnings estimates for female workers younger than age 60. Chart 20 reveals much more variation between the estimates for female workers aged 60 or older, mirroring the pattern of divergence seen for older men.

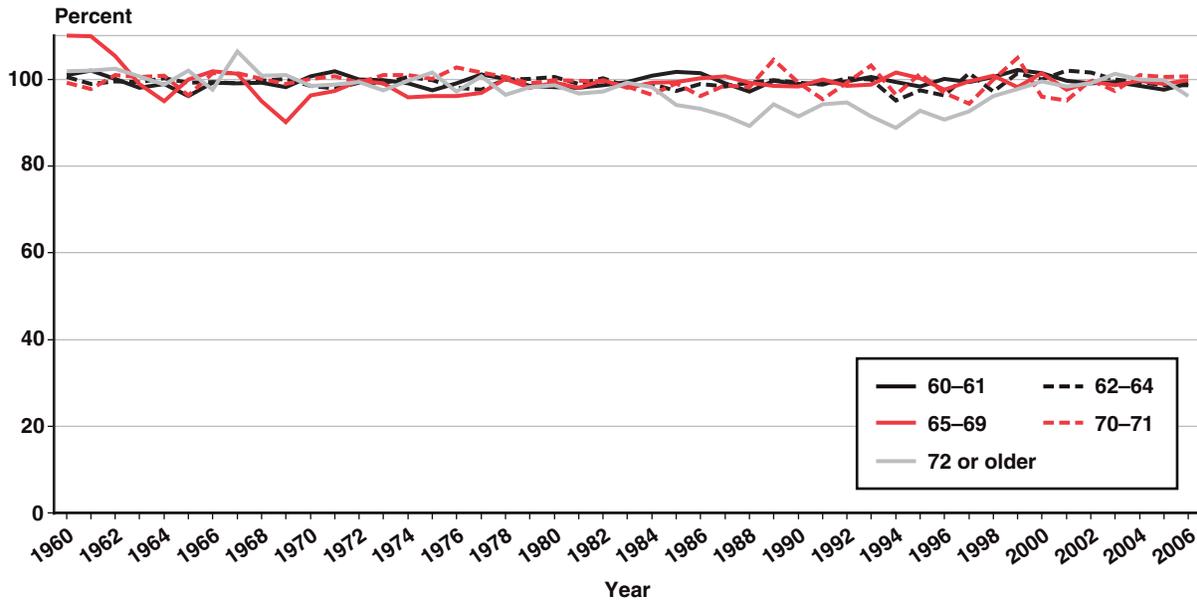
Table 7.
Supplement estimates of median earnings for men as a percentage of the taxable maximum amount, by age group, 1960–1980

Year	19 or younger	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–61	62–64	65–69	70–71	72 or older
1960	12.81	44.08	82.13	100.65	107.81	107.65	103.21	97.96	92.00	82.96	79.42	44.00	25.15	27.92
1961	12.96	43.29	82.42	102.56	110.42	110.81	107.77	100.90	94.69	84.90	82.60	44.17	25.58	28.60
1962	13.46	44.19	87.19	107.38	116.25	118.06	113.44	106.33	100.77	92.08	85.98	46.73	24.75	29.17
1963	12.75	46.08	90.98	111.75	120.42	122.44	117.94	110.77	104.54	97.23	89.56	48.10	24.77	28.79
1964	13.44	49.83	97.13	117.67	126.06	128.60	126.44	117.31	110.69	101.06	93.83	52.52	24.79	29.75
1965	14.79	54.35	103.27	126.15	135.02	135.81	132.69	123.35	116.27	104.02	99.67	54.75	25.96	30.06
1966	12.44	40.56	82.58	97.32	104.56	106.42	104.56	98.03	90.95	83.88	76.61	36.14	21.92	23.47
1967	12.09	38.97	86.53	102.05	108.38	110.14	109.12	102.94	94.71	88.30	80.71	37.82	22.29	23.36
1968	10.87	35.74	78.90	93.51	99.37	101.35	100.33	94.73	86.41	80.82	72.51	33.91	20.53	21.55
1969	11.67	39.33	85.26	100.08	105.60	108.10	107.33	101.95	93.95	86.99	77.94	35.96	21.18	22.62
1970	11.92	42.06	87.53	104.24	109.33	111.36	112.22	107.31	98.40	90.40	82.77	37.53	21.31	23.88
1971	12.05	31.62	90.68	108.50	112.55	118.35	117.22	112.79	104.71	95.36	86.83	38.10	21.40	24.55
1972	11.47	45.68	82.28	102.89	109.22	111.33	110.67	107.44	99.80	90.03	79.64	32.26	18.69	22.34
1973	10.82	42.01	72.94	93.61	99.35	102.13	101.94	99.47	91.68	83.19	71.55	25.91	18.06	20.52
1974	9.66	36.88	62.85	81.45	87.47	90.14	90.27	87.70	80.91	73.86	62.61	21.33	16.48	17.61
1975	8.84	34.54	60.03	79.22	86.04	88.89	89.40	87.02	80.07	73.74	61.70	20.53	16.14	16.82
1976	8.90	33.99	58.92	78.59	86.46	89.05	90.27	87.68	81.44	73.41	62.10	19.37	15.92	16.94
1977	8.86	34.09	57.73	77.35	86.85	89.36	90.13	88.12	82.27	74.32	62.32	19.02	15.78	16.52
1978	9.28	34.95	58.55	77.24	88.82	90.96	92.20	89.79	84.54	76.71	63.15	21.89	16.86	17.39
1979	7.96	29.66	49.37	64.23	74.17	76.72	77.08	75.45	71.27	64.76	53.63	19.40	14.66	14.52
1980	7.17	27.05	45.87	59.81	69.08	72.28	72.96	71.01	67.90	61.54	50.97	18.93	14.12	13.63

SOURCE: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, various editions.

Chart 18.

EPUF estimates of median earnings as a percentage of the Supplement estimate, male workers aged 60 or older by age group, 1960–2006

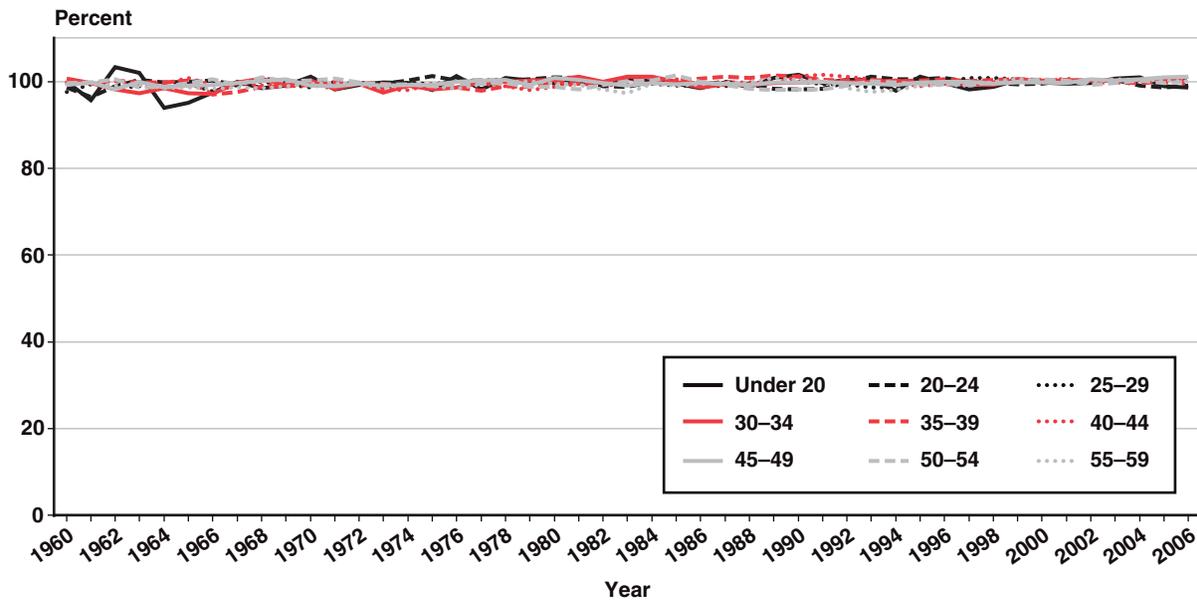


SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B6, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 19.

EPUF estimates of median earnings as a percentage of the Supplement estimate, female workers younger than age 60 by age group, 1960–2006

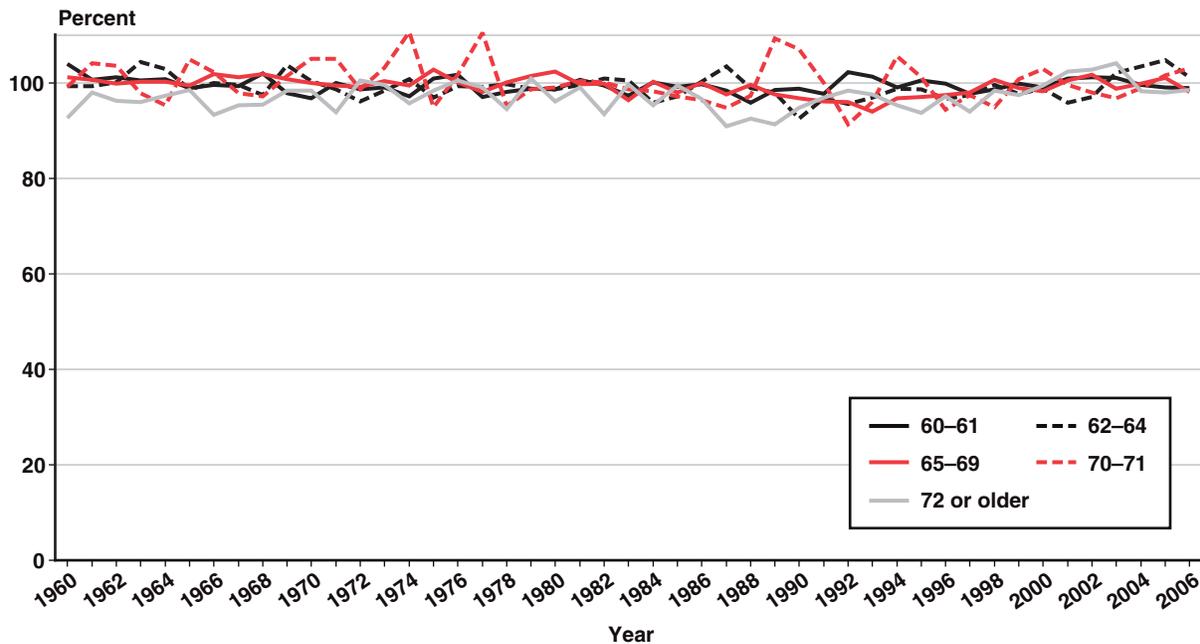


SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B6, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Chart 20.

EPUF estimates of median earnings as a percentage of the Supplement estimate, female workers aged 60 or older by age group, 1960–2006



SOURCES: SSA, *Annual Statistical Supplement to the Social Security Bulletin*, Table 4.B6, various editions; and author's calculations using EPUF.

NOTE: 2008 Supplement estimates for 2004–2006 were based on preliminary data.

Percentage of Workers with Earnings Below the Taxable Maximum

Supplement Table 4.B4 contains estimates of the percentage of all workers with earnings below the taxable maximum amount beginning in 1951. Table 8 compares the EPUF estimates with those found in the Supplement. Few of the estimates differ by more than one-tenth of a percentage point. For all workers, the largest difference between the estimates occurs in 1967, where the EPUF estimate is 0.6 percent greater than the Supplement estimate. The largest difference in the estimates of workers by sex is seen for 1969, in which EPUF estimates are 1.7 percent higher for men and 1.0 percent lower for women than the Supplement estimates. Apart from these and scattered other modest differences, the EPUF and Supplement estimates scarcely differ.

Summary

This analysis compares the earnings data contained in EPUF with estimates published in the *Annual Statistical Supplement to the Social Security Bulletin*. The analysis presents four reasons why one should expect differences between the estimates beyond those due to the different sampling frames used to generate the respective 1-percent samples.

First, the Supplement estimates are based on taxable earnings, which can sum to more than the Social Security taxable maximum for multiple jobholders, whereas the EPUF reflects only capped taxable earnings. Second, EPUF data cleaning and disclosure prevention measures reduce the number of records with earnings and the amount of earnings reported. Third, the Supplement estimates reflect adjustments to account for delinquent, erroneous, and potentially fraudulent reporting of earnings information to

Table 8.**Comparing Supplement and EPUF estimates: Percentage of all, male, and female workers with earnings below the taxable maximum amount, 1951–2006**

Year	Supplement ^a			EPUF			EPUF estimate as a percentage of Supplement estimate		
	All	Men	Women	All	Men	Women	All	Men	Women
1951	75.5	64.6	96.7	75.4	64.7	96.6	99.9	100.1	99.9
1952	72.1	60.0	95.4	72.2	60.1	95.3	100.2	100.2	99.9
1953	68.8	55.5	93.8	68.8	55.5	93.8	100.0	100.0	100.0
1954	68.4	55.4	93.0	68.4	55.4	93.1	100.0	100.0	100.1
1955	74.4	63.4	95.9	74.7	63.8	95.8	100.4	100.7	99.9
1956	71.6	59.7	94.5	71.8	59.9	94.6	100.2	100.4	100.1
1957	70.1	58.7	93.1	70.2	58.8	93.0	100.1	100.2	99.9
1958	69.4	58.4	91.8	69.6	58.6	91.7	100.2	100.3	99.9
1959	73.3	62.7	94.3	73.3	62.6	94.2	99.9	99.9	99.9
1960	72.0	60.9	93.5	71.9	60.7	93.4	99.8	99.7	99.9
1961	70.8	59.6	92.4	70.8	59.6	92.4	100.0	99.9	100.0
1962	68.8	57.1	91.1	68.8	57.0	91.1	100.0	99.9	100.0
1963	67.5	55.5	90.0	67.4	55.4	90.0	99.9	99.8	100.0
1964	65.5	53.1	88.5	65.5	53.0	88.5	100.0	99.8	100.0
1965	63.9	51.0	87.3	63.9	50.8	87.4	100.0	99.7	100.1
1966	75.8	64.4	95.6	75.7	64.2	95.7	99.9	99.7	100.1
1967	73.6	61.5	94.2	74.0	62.0	94.5	100.6	100.8	100.3
1968	78.6	68.0	96.3	78.6	67.9	96.3	100.0	99.9	100.0
1969	75.5	62.8	96.0	75.8	63.9	95.0	100.4	101.7	99.0
1970	74.0	61.8	93.5	73.9	61.6	93.5	99.8	99.7	100.0
1971	71.7	59.1	91.7	71.6	58.9	91.8	99.9	99.6	100.1
1972	75.0	62.9	93.9	74.9	62.7	93.8	99.9	99.7	99.9
1973	79.7	68.9	96.2	79.6	68.6	96.2	99.9	99.6	100.0
1974	84.9	76.2	97.8	84.8	76.0	97.8	99.9	99.7	100.0
1975	84.9	76.4	97.5	85.0	76.4	97.5	100.1	100.0	100.1
1976	85.1	76.3	97.5	85.0	76.2	97.5	99.9	99.9	100.0
1977	85.2	76.3	97.5	85.2	76.3	97.5	99.9	99.9	100.0
1978	84.6	75.4	97.1	84.6	75.4	97.3	100.0	99.9	100.2
1979	90.0	83.6	98.6	90.1	83.7	98.8	100.2	100.1	100.2
1980	91.2	85.5	98.8	91.3	85.5	98.9	100.1	100.0	100.1
1981	92.4	87.4	99.0	92.4	87.3	99.0	100.0	99.9	100.0
1982	92.9	88.3	98.9	92.9	88.2	99.0	100.0	99.9	100.1
1983	93.7	89.6	99.0	93.7	89.6	99.0	100.0	100.0	100.0
1984	93.6	89.4	98.9	93.6	89.4	98.9	100.0	100.0	100.0
1985	93.5	89.3	98.8	93.5	89.3	98.8	100.0	100.0	100.0
1986	93.8	89.7	98.7	93.8	89.8	98.7	100.0	100.1	100.0
1987	93.9	89.9	98.6	93.8	89.9	98.6	99.9	100.0	100.0
1988	93.5	89.4	98.3	93.5	89.4	98.4	100.0	100.0	100.1
1989	93.8	90.1	98.3	93.8	90.0	98.3	100.0	99.9	100.0
1990	94.3	90.9	98.4	94.3	90.9	98.4	100.0	100.0	100.0
1991	94.4	91.1	98.3	94.4	91.1	98.3	100.0	100.0	100.0
1992	94.3	91.0	98.1	94.3	91.0	98.1	100.0	100.0	100.0
1993	94.4	91.3	98.1	94.4	91.3	98.1	100.0	100.0	100.0
1994	94.6	91.4	98.1	94.6	91.5	98.1	100.0	100.1	100.0
1995	94.2	91.0	97.9	94.2	91.0	97.9	100.0	99.9	100.0

(Continued)

Table 8.**Comparing Supplement and EPUF estimates: Percentage of all, male, and female workers with earnings below the taxable maximum amount, 1951–2006—Continued**

Year	Supplement ^a			EPUF			EPUF estimate as a percentage of Supplement estimate		
	All	Men	Women	All	Men	Women	All	Men	Women
1996	93.9	90.6	97.7	93.9	90.6	97.7	100.0	100.0	100.0
1997	93.8	90.5	97.6	93.8	90.5	97.6	100.0	100.0	100.0
1998	93.7	90.3	97.5	93.7	90.3	97.5	100.0	100.0	100.0
1999	93.9	90.7	97.5	93.9	90.7	97.5	100.0	100.0	100.0
2000	93.8	90.6	97.4	93.8	90.6	97.4	100.0	100.0	100.0
2001	94.1	91.0	97.5	94.0	91.0	97.4	99.9	100.0	99.9
2002	94.6	91.8	97.7	94.5	91.7	97.6	99.9	99.9	99.9
2003	94.5	91.8	97.5	94.5	91.7	97.5	100.0	99.9	100.0
2004	^b 94.1	^b 91.2	^b 97.3	94.1	91.2	97.3	100.0	100.0	100.0
2005	^b 93.9	^b 91.0	^b 97.1	93.9	90.9	97.1	100.0	99.9	100.0
2006	^b 94.0	^b 91.1	^b 97.1	93.9	91.0	97.1	99.9	99.8	100.0

SOURCES: SSA (2009, Table 4.B4) and author's calculations using EPUF.

a. From 1937 to 1950, relates to wage and salary workers. Beginning in 1951, includes self-employed workers.

b. 2008 Supplement estimates were based on preliminary data.

SSA. Fourth, the Supplement updates only the three most recent years of estimates. As a result, older estimates are frozen and do not reflect any subsequent changes in the MEF. EPUF earnings data reflect the continuously updated MEF and contain the most recent earnings data reported to SSA.

The analysis began by comparing estimates of taxable earnings in the underlying EPUF sample with those in the Supplement. Those estimates proved very similar and supported the expectation that the biggest differences would be for the most recent years. Although there was some divergence between the estimates from the underlying EPUF sample and the Supplement estimates from 1980 through 1988, those differences were relatively minor.

In general, the other differences between EPUF and Supplement earnings estimates are relatively small after accounting for expected differences. The key differences are largely attributable to EPUF's use of capped taxable earnings and the removal of some records due to data cleaning and disclosure prevention procedures.

There were, however, two unexpected differences between EPUF and Supplement estimates: Specifically, estimates of the number of workers by age group and sex, and the value of median earnings by age and sex. The distribution of individuals who have earnings when they are 72 years old or older in EPUF is nearly identical to that in the active file of the CWHS, indicating that EPUF represents the current state of the earnings data contained in the MEF, even though it differs from the Supplement estimates. Median earnings estimates differ because Supplement estimates are adjusted to account for estimated earnings above the taxable maximum and EPUF estimates are not.

Appendix

EPUF's two linkable subfiles differ in structure and in the ways the data cleaning and disclosure protection procedures affect them. The demographic and aggregate earnings subfile uses a person-record format containing a single record for each of the 4,348,254 individuals in the EPUF. Each record contains the individual's EPUF identification code, year of birth, sex, aggregate taxable earnings from 1937 through 1950, aggregate quarters of coverage earned from 1937 through 1950, and quarters of coverage earned in 1951 and 1952. Appendix Table A-2 presents an illustrative listing of 20 hypothetical demographic and aggregate earnings subfile records. The annual earnings subfile is a vertical-event history file that contains a single record for each year with positive earnings for each person in EPUF. Every earnings-year record contains the individual's EPUF identification code, capped taxable earnings, and earned quarters of coverage. The annual earnings subfile contains 60,326,474 records for the 3,131,424 individuals who had at least 1 year of positive earnings from 1951 through 2006. Appendix Table A-3 presents an illustrative listing of 41 earnings years for four hypothetical earners.

Nearly 28 percent of the individuals in EPUF had no positive annual earnings from 1951 through 2006. These individuals have a record in the demographic and aggregate earnings subfile, but no records in the annual earnings subfile.

The MEF 1-percent sample underlying the EPUF contained records for 4,413,024 individuals. Data "cleaning" led to the removal of records for 28,770 individuals; those records had dubious or missing year-of-birth values, coding errors, or other issues. Then, to protect against disclosure of personal data, earnings records for individuals aged 14 or younger or 86 or older were zeroed out, effectively removing those records from the annual earnings subfile (because the subfile only contains records with positive earnings values). However, setting those earnings equal to \$0 does not affect the individual's record in the demographic and aggregate earnings subfile.

Table A-1.
Taxable earnings amounts removed from underlying EPUF sample by reason for removal, 1951–2006

Year	Total taxable earnings in underlying EPUF sample (million \$)	Capping (earnings exceeding the taxable maximum)		Data cleaning and disclosure prevention		Total earnings removed from underlying EPUF sample	
		Dollars (in millions)	As a percentage of taxable earnings	Dollars (in millions)	As a percentage of taxable earnings	Dollars (in millions)	As a percentage of taxable earnings
1951	1,206.9	24.8	2.06	5.7	0.47	30.5	2.53
1952	1,287.8	31.3	2.43	5.9	0.46	37.2	2.89
1953	1,357.9	36.3	2.67	6.0	0.44	42.3	3.11
1954	1,334.2	32.7	2.45	5.9	0.44	38.5	2.89
1955	1,570.5	30.2	1.92	7.0	0.44	37.2	2.37
1956	1,705.6	38.3	2.25	7.5	0.44	45.8	2.69
1957	1,818.0	43.0	2.37	7.6	0.42	50.6	2.78
1958	1,801.3	40.6	2.25	7.4	0.41	48.0	2.67
1959	2,014.9	41.2	2.04	8.1	0.40	49.3	2.45
1960	2,067.3	44.0	2.13	8.2	0.40	52.1	2.52
1961	2,091.5	45.3	2.17	8.1	0.39	53.4	2.55
1962	2,187.2	53.4	2.44	8.3	0.38	61.7	2.82
1963	2,252.2	57.4	2.55	8.5	0.38	65.9	2.92
1964	2,359.8	66.9	2.84	8.8	0.37	75.7	3.21
1965	2,498.0	79.1	3.17	9.2	0.37	88.3	3.54

(Continued)

Table A-1.**Taxable earnings amounts removed from underlying EPUF sample by reason for removal, 1951–2006—Continued**

Year	Total taxable earnings in underlying EPUF sample (million \$)	Capping (earnings exceeding the taxable maximum)		Data cleaning and disclosure prevention		Total earnings removed from underlying EPUF sample	
		Dollars (in millions)	As a percentage of taxable earnings	Dollars (in millions)	As a percentage of taxable earnings	Dollars (in millions)	As a percentage of taxable earnings
1966	3,116.6	63.6	2.04	11.3	0.36	74.9	2.40
1967	3,274.5	73.4	2.24	11.6	0.36	85.0	2.60
1968	3,745.5	68.5	1.83	13.1	0.35	81.6	2.18
1969	4,014.2	89.2	2.22	13.7	0.34	102.9	2.56
1970	4,142.5	95.2	2.30	13.8	0.33	108.9	2.63
1971	4,255.7	101.1	2.38	14.1	0.33	115.2	2.71
1972	4,824.7	99.5	2.06	15.6	0.32	115.2	2.39
1973	5,591.1	91.4	1.63	17.6	0.32	109.0	1.95
1974	6,340.4	74.4	1.17	19.7	0.31	94.0	1.48
1975	6,631.7	70.9	1.07	20.1	0.30	91.0	1.37
1976	7,351.2	78.8	1.07	21.5	0.29	100.3	1.37
1977	8,121.2	87.0	1.07	22.8	0.28	109.8	1.35
1978	9,126.4	122.7	1.34	24.3	0.27	147.0	1.61
1979	10,662.1	93.7	0.88	27.3	0.26	121.0	1.13
1980	11,680.1	92.1	0.79	28.2	0.24	120.3	1.03
1981	12,898.0	89.7	0.70	28.2	0.22	118.0	0.91
1982	13,533.3	85.8	0.63	28.0	0.21	113.8	0.84
1983	14,400.7	80.5	0.56	28.3	0.20	108.9	0.76
1984	15,829.2	96.0	0.61	29.2	0.18	125.2	0.79
1985	17,059.9	105.7	0.62	30.0	0.18	135.7	0.80
1986	18,180.7	108.5	0.60	30.4	0.17	138.9	0.76
1987	19,391.7	114.6	0.59	30.3	0.16	144.8	0.75
1988	20,829.8	130.6	0.63	30.2	0.14	160.8	0.77
1989	22,249.7	135.5	0.61	30.7	0.14	166.2	0.75
1990	23,451.3	130.9	0.56	30.8	0.13	161.7	0.69
1991	24,078.2	130.3	0.54	29.8	0.12	160.0	0.67
1992	25,184.5	146.3	0.58	30.6	0.12	177.0	0.70
1993	26,179.2	158.5	0.61	31.2	0.12	189.8	0.72
1994	27,701.5	182.0	0.66	32.2	0.12	214.2	0.77
1995	29,020.8	202.9	0.70	33.0	0.11	235.9	0.81
1996	30,564.4	239.0	0.78	33.9	0.11	272.8	0.89
1997	32,654.2	271.4	0.83	35.5	0.11	306.9	0.94
1998	34,991.9	303.9	0.87	37.3	0.11	341.2	0.98
1999	37,156.0	318.3	0.86	38.9	0.10	357.3	0.96
2000	39,651.7	398.2	1.00	40.2	0.10	438.4	1.11
2001	41,167.3	345.0	0.84	40.5	0.10	385.5	0.94
2002	41,919.1	283.0	0.68	40.1	0.10	323.1	0.77
2003	42,922.0	275.9	0.64	39.7	0.09	315.6	0.74
2004	44,773.0	319.6	0.71	40.7	0.09	360.3	0.80
2005	46,539.3	358.1	0.77	40.5	0.09	398.6	0.86
2006	48,843.1	411.4	0.84	40.4	0.08	451.8	0.93
Total	871,600.1	7,387.7	0.85	1,267.6	0.15	8,655.3	0.99

SOURCE: Author's calculations using underlying EPUF sample.

Table A-2.
Illustrative examples of person records in the EPUF Demographic and Aggregate Earnings subfile

ID number	Year of birth (YOB)	Sex ^a	Total covered earnings (\$) (TOT_COV_EARN3750)	Quarters of coverage 1937–1950 (QC3750)	Quarters of coverage 1951–1952 (QC5152)
1	1973	1	0	0	0
2	1976	2	0	0	0
3	1917	2	9,300	18	3
4	1947	2	0	0	0
5	1983	1	0	0	0
6	1927	2	0	0	0
7	1995	2	0	0	0
8	1996	2	0	0	0
9	1931	1	39	0	2
10	1984	2	0	0	0
11	1961	2	0	0	0
12	1983	2	0	0	0
13	1914	2	13,100	26	0
14	1918	1	9,400	18	7
15	1932	2	225	2	0
16	1978	2	0	0	0
17	1937	1	0	0	0
18	1945	2	0	0	0
19	1985	1	0	0	0
20	1921	1	15,900	31	6

SOURCE: Author's reconstruction based on EPUF.

a. 1 = male, 2 = female.

Table A-3.
Illustrative examples of earnings-year records in the EPUF Annual Earnings subfile

ID number	Year with earnings (YEAR_EARN)	Quarters of coverage (ANNUAL_QTRS)	Capped taxable earnings (\$) (ANNUAL_EARNINGS)
1	1998	4	7,500
1	1999	4	10,000
1	2000	4	16,200
1	2001	4	24,000
1	2002	4	15,900
1	2004	3	3,600
1	2005	4	20,500
1	2006	4	24,300
2	1993	2	1,600
2	1994	4	2,600
2	1995	3	2,100
2	1996	4	4,500
2	1997	4	7,600
2	1998	4	22,700
2	1999	4	16,900
2	2000	4	26,300
2	2001	4	33,200
2	2002	4	36,300
2	2003	4	41,900
2	2004	4	42,100
2	2005	4	38,500
2	2006	4	29,800
3	1951	0	550
3	1952	0	325
3	1953	4	575
3	1954	3	2,000
3	1955	3	1,600
4	1966	1	675
4	1967	4	3,200
4	1968	4	4,000
4	1969	4	4,400
4	1970	4	4,800
4	1971	4	5,000
4	1972	4	5,300
4	1973	4	5,800
4	1974	4	6,300
4	1975	4	7,200
4	1976	4	8,000
4	1977	4	8,600
4	1978	4	6,600
4	1980	4	3,000

SOURCE: Author's reconstruction based on EPUF.

References

- Buckler, Warren. 1988. "Commentary: Continuous Work History Sample." *Social Security Bulletin* 51(4): 12, 56.
- Compson, Michael. 2011. "The 2006 Earnings Public-Use Microdata File: An Introduction." *Social Security Bulletin* 71(4): 33–59.
- Olsen, Anya, and Russell Hudson. 2009. "Social Security Administration's Master Earnings File: Background Information." *Social Security Bulletin* 69(3): 29–45.
- Smith, Creston M. 1989. "The Social Security Administration's Continuous Work History Sample." *Social Security Bulletin* 52(10): 20–28.
- [SSA] Social Security Administration. 2002. *Congressional Response Report: Status of the Social Security Administration's Earnings Suspense File*. Report No. A-03-03-23038. Baltimore, MD: SSA, Office of the Inspector General.
- . 2009. *Annual Statistical Supplement to the Social Security Bulletin, 2008*. Washington, DC: SSA, ORES.
- . 2011. *Annual Statistical Supplement to the Social Security Bulletin, 2010*. Washington, DC: SSA, ORES.