

Evaluating Respondents' Reporting of Social Security Income In the Survey of Income and Program Participation (SIPP) Using Administrative Data

Lydia Scoon-Rogers¹

**U.S. Bureau of the Census
HHES Division, F.O.B. #3,
LIS Branch, Rm. 1425-3 LIS Branch, Rm. 1425-3, Washington, D.C. 20233-8500, Lydia.Megan.Scoon.Rogers@census.gov**

Introduction

In the 1996 panel of the Survey of Income and Program Participation (SIPP), respondents were asked to exclude Medicare Part B from the amount of before-tax social security income they received. Medicare Part B, also known as Supplementary Medical Insurance (SMI), is health insurance that most Medicare beneficiaries have the option to buy or in some circumstances receive at no cost. SMI helps pay for doctor's fees, outpatient visits, medical services, and supplies not covered under Medicare Part A. The majority of Medicare beneficiaries who elect to receive SMI have a premium that is deducted from their monthly social security income benefits. It was not collected in the 1996 SIPP panel because it was thought that people who pay for it might be unaware of the amount that gets deducted from their monthly social security benefit check. The 2001 and 2004 panels created a means to further explore if this amount of before-tax income was salient by asking about SMI amounts.² This paper has two goals: first, to identify and correct for any SMI and other types of errors found in the 1996 SIPP by linking the SIPP data to Social Security Administration (SSA) data and comparing the results. The second goal is to add missing SMI premiums to 1996 SIPP panel social security income data, and see how this affects the poverty status of those 65 years old and over.

¹

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

²Part of the 2001 SIPP panel asked respondents to include SMI in the one question about social security amounts. The 2004 panel asks respondents to report their social security benefit as they know it, then asks follow-up questions about whether that amount included SMI or not. Additional analyses as are found in this paper are underway concerning these more recent SIPP panels.

Methodology

The above issues were investigated using a special research file, created in cooperation with the SSA that matched person records from the 1996 SIPP panel file by social security number to the SSA Payment History Update System (PHUS data).³ The SIPP is a U.S. Census Bureau longitudinal survey of civilian, noninstitutionalized households. It's conducted at 4-month intervals and provides detailed monthly data about income, poverty and other social and demographic topics. In SIPP, questions are asked about how much income is received each month for the prior 4 months. The SSA's PHUS data file is a record of what was paid to social security beneficiaries monthly, reflecting all adjustments for any double checks sent, death, or for other administrative accounting.

As has been the case in earlier studies, amounts reported in the administrative data provide a benchmark to assess the estimates reported in SIPP. That's because identified errors have been found to be of a lower magnitude in the SSA records than in the SIPP data (Huynh, Rupp, Sears, 2001). Potential sources of error from the SSA file include incorrect addresses, lost checks and other unexpected accounting errors. Potential sources of error in SIPP come from type z responses,⁴ proxy responses, imputation, rounding errors and under- or over-reporting of amounts. They may also stem from a mis-classification with other income sources, such as Supplemental Security Income (SSI), Supplemental Security Disability Insurance (SSDI), and other disability income or private pensions. In addition, there is also the potential for reported SIPP social security amounts to reflect the combining of benefits from other family members, such as a spouse or eligible dependent children. (Vaughan, 2003).

The universe for analysis was people age 65 years and older with a positive social security income amount reported in either the SIPP and SSA. Cross-sectional, weighted person records from the 1996 SIPP panel for March 1997, 1998 and 1999 were used. In March 1999, the number of people age 65 and older with positive SIPP and/or SSA benefit amounts totaled approximately 26.2 million⁵ (unweighted N=6,047).⁶ This comprised 81.0 percent of all SIPP elderly (matched or unmatched), and 95.5 percent of all matched SIPP elderly. The match rate for the elderly was 84.7 percent, similar to SIPP-SSA match rates cited by other researchers for this SIPP panel (Table 1).

According to the PHUS data, 87.4 percent of the 26.2 million elderly beneficiaries paid for the SMI coverage. Some of the 12.6 percent that did not pay chose not to or were ineligible to participate in the SMI program. Others that did not pay got SMI coverage free. The SMI premium cost in 1999 was \$45.50 for 97.6 percent of SMI payees and did not differ from that by more than \$50 for the remaining payees.

³In order to protect confidentiality, social security numbers were replaced with a unique person identifier.

⁴Type z responses are where one or more [but not all] persons in the household were not interviewed because they refused or were away at the time of the interview or had moved out of the surveyed household, and a proxy interview was not obtained. Data are imputed for these persons in households.

⁵ Among these elderly people with a positive dollar amount in one or both sources,) 0.3 million had a positive social security amount in the PHUS, and a zero benefit amount in SIPP. Some of these are likely false negatives meaning people actually received social security as the PHUS data indicate, but they did not report it in SIPP or mis-classified it as another income source, such as SSI or a pension. Conversely, SIPP identified 0.4 million beneficiaries with positive social security amounts and zero SSA benefit amounts. Some of these are likely false positives in SIPP where an income source other than social security was classified as social security income in SIPP. More research needs to be done to investigate sources of these discrepancies.

⁶For broad contextual purposes, the paper shows all results in terms of the estimated population of elderly social security beneficiaries, weighted, instead of in terms of the actual number in sample (N=6,047 persons with positive social security amounts in SIPP or SSA).

Table 1. Poverty Rates for SIPP Adults Whose Data Are Matched to Administrative Data, by Match Status and Adjustment to Social Security Income							
				Poverty Rate		Number (In Thousands)	
				15 Years and Older	65 Years and Older	15 Years and Older	65 Years and Older
Total (SIPP estimate, no adjustment)				11.4	10.8	211,598	32,385
Not matched to the PHUS data (SIPP estimate, no adjustment)				10.5	14.7	146,925	4,941
Matched cases only (SIPP estimate, no adjustment)				13.3	10.1	64,673	27,443
With \$1+ social security in either SIPP or PHUS (SIPP estimate, no adjustment)				11.8	9.8	35,659	26,221
PHUS direct pay replaced SIPP (for error)				10.5	8.5	35,659	26,221
PHUS direct pay and PHUS SMI replaced (for SMI omission)				9.6	7.5	35,659	26,221
Note: Poverty rate calculated by using monthly total family income and official annual poverty thresholds, the latter divided by 12.							
Source: Survey of Income and Program Participation (SIPP) 1996 Panel Wave 10 File and linked data from the Social Security Administration's Payment History Update System (PHUS).							

The March 1999 mean social security amount reported in SIPP was \$691 compared with \$724 in the PHUS file. The distributions of social security income from the two sources varied somewhat; slightly more than one fourth (28.6 percent) of elderly recipients had benefit amounts under \$500 according to the SSA file and a little more than that (30.8 percent) had benefits at that level according to the SIPP file – mainly in the \$300 to \$499 range for both sources. Almost one-half of beneficiaries received \$500 to \$900 according to each data source (43.7 percent for the SSA and a comparable 44.6 percent for the SIPP). The remaining beneficiaries received amounts of \$900 to approximately \$7,000, with the SSA slightly more likely to have reported these higher levels than SIPP (27.6 percent for SSA and 24.5 percent for SIPP) (Table 2). ⁷

⁷ The proportion of social security beneficiaries correctly reporting social security amounts in 1998 (39.6 percent) and the proportion that showed SIPP amounts lower than SSA amounts (41.8 percent) were not statistically different.

Table 2. Distributions of Monthly Social Security Income for Beneficiaries 65 and Older from SIPP and Matched Administrative Data: March 1997, March 1998, and March 1999 (Numbers in Thousands)

[illegible]

Findings

Toward answering how well SIPP beneficiaries fared in excluding the SMI amount, as directed in the 1996 panel, 37.7 percent of SIPP beneficiaries in 1999 correctly omitted the Medicare Part B amount from their social security benefits.⁸ This was lower than the proportion of beneficiaries that correctly reported social security income in March 1998 and 1997 (39.6 percent in 1998, comparable with 42.1 percent in 1997). Correct reporting was ascertained because SIPP beneficiaries reported receiving a monthly social security amount that exactly matched the SSA amount, excluding the SMI. About 42.5 percent of SIPP beneficiaries reported an amount lower than the SSA amount in 1999, comparable with 41.8 percent in 1998 and 40.8 percent in 1997 that reported lower SIPP amounts than actually received. The remaining 19.8 percent of the 1999 SIPP beneficiaries reported an amount greater than the administrative data, comparable with the 18.6 percent in 1998 but a little higher than the 17.1 percent reported in 1997 (Table 3).

Table 3. Relative Levels of Monthly Social Security Income from SIPP and Matched Administrative Data for Beneficiaries 65 Years and Older: March 1997, 1998, and 1999				
	March 1997	March 1998	March 1999	
Total (Numbers in thousands)	25,788	25,900	26,221	
Equal, nonzero amounts	10,861	10,244	9,882	
Administrative > SIPP	10,521	10,838	11,145	
SIPP > Administrative	4,406	4,818	5,194	
Percent	100.0	100.0	100.0	
Equal, nonzero amounts	42.1	39.6	37.7	
Administrative > SIPP	40.8	41.8	42.5	
SIPP > Administrative	17.1	18.6	19.8	
Universe is persons 65+ with positive social security benefits in SIPP and/or SSA.				
Source: Survey of Income and Program Participation (SIPP) 1996 panel wave files linked data from the Social Security Administration (SSA)				

⁸This may include some or all of those elderly who did not pay an SMI premium.

In sum, 37.7 percent of SIPP elderly social security beneficiaries correctly excluded their SMI amount in 1999. The status of SMI reporting for the remaining 62.3 percent of the elderly was less clear and required further inquiry. The next step was to look at whether other sources of error (non-SMI error) might be masking SIPP's exclusion of the SMI amount for any of them. The types of non-SMI error investigated for 1999 were cost of living adjustment error, imprecision, and rounding.

Cost of living adjustment error

One of the smallest magnitudes of error would be respondents not reporting their cost of living adjustments (COLAs) in their monthly benefit amount. The COLA adjustment for 1999 social security benefits was 1.3 percent and started in December 1998 but was *first received* in January 1999.⁹ Since the social security benefits in this report reflect what people received early in the calendar year (March), it seems plausible that some SIPP beneficiaries may have overlooked reporting the newly adjusted amount first *received* in January and continued this mis-reporting through March or later. After comparing March 1999 SIPP social security amounts with SSA amounts received in December 1998, 5.4 percent (1.4 million) SIPP elderly beneficiaries erroneously reported the December monthly amount. These beneficiaries appeared to have correctly omitted SMI amounts. (Table 4)

Imprecision error

It may be reasonable to expect that respondents made errors in precision. That is, respondents may not have recalled the exact amount of social security income received. It could be that some error in social security amounts was due to proxy responding, or the respondent did not use any records and could not recall the exact amount. It's probably reasonable to assume that some of the nominal difference between reported SIPP amounts and the amounts in the SSA records is due to sheer imprecision on the part of the SIPP respondent. Two views of imprecision were evaluated -- small dollar differences and small percent differences.

Nominal dollar differences. One-fourth (24.7 percent) of elderly social security beneficiaries had a SSA-SIPP absolute value benefit differential of under \$25 in 1999. As much as 13.6 percent of beneficiaries reported an error level of up to \$10. It is quite feasible to think that where the difference between the SIPP and SSA amounts was small, (i.e. under \$25) that the Medicare Part B could also have been properly excluded. (Table 4)

Nominal percentage differences. A nominal percentage difference was defined as the SIPP amount differing from the SSA amount by less than 5 percent. This definition tended to identify the same beneficiaries as did identifying people with less than a \$25 difference. Specifically, almost all of the 6.5 million SIPP beneficiaries who showed a SIPP-SSA differential of under \$25 in 1999 also had a percent differential less than 5 percent of the total SSA social security amount. The fact that the purported precision error also comprised only a small share of SIPP social security monthly income would suggest that it's likely an inadvertent error. Also, some might guess that imprecision would largely be random, not a characteristic of any one socioeconomic group. One such finding was that people with low incomes were as likely to have a SIPP-SSA differential of under \$25 as those people with high incomes.¹⁰ (Table 4)

⁹ SSA's press office confirmed that social security cost of living increases announced for the upcoming year (i.e. for 1999) start in December of the prior year (i.e. December 1998). For example, new amounts are received in January 1999, representing the December increase. Therefore, the last month that the social security recipient receives the "old, prior year amount" is December 1998, representing November's payment.

¹⁰ An interesting side note observed in Table 5 is that at the upper end of the SIPP-SSA absolute value differential distribution, there are differences in likelihood of poor and nonpoor having a SIPP-SSA gap of more than \$150. The fact that the poor are more likely to show such a difference may, to some degree, reflect some SIPP mis-reporting of social security income with SSI, the latter which is a government benefit for the low-income elderly and disabled.

Table 4. Absolute Value Difference Between Administrative and SIPP Monthly Social Security Income for SIPP Beneficiaries 65 Years and Older with \$1 or More Difference: March 1997, 1998, and 1999

		March 1997	March 1998	March 1999	March 1997	March 1998	March 1999
		Numbers			Percent		
Total (Number in thousands)		25,788	25,900	26,221	100.0	100.0	100.0
Nonzero difference (SSA > or < SIPP)							
	Total	14,925	15,656	16,339	57.9	60.4	62.3
	\$1 to 24	5,748	5,990	6,469	22.3	23.1	24.7
	\$1-4	1,451	1,634	1,845	5.6	6.3	7.0
	\$5-9	1,148	1,194	1,739	4.5	4.6	6.6
	\$10-14	1,128	1,105	1,386	4.4	4.3	5.3
	\$15-19	980	1,206	778	3.8	4.7	3.0
	\$20-24	1,041	851	721	4.0	3.3	2.7
	\$25 to 49	2,746	2,761	2,663	10.6	10.7	10.2
	\$50 to 74	1,108	1,310	1,235	4.3	5.1	4.7
	\$75 to 149	1,813	1,658	1,873	7.0	6.4	7.1
	\$150-499	2,521	2,673	2,810	9.8	10.3	10.7
	\$500+	989	1,264	1,289	3.8	4.9	4.9
Total Nonpoor (Number in thousands)		23,139	23,395	23,657	100.0	100.0	100.0
Nonzero difference (SSA > or < SIPP)							
	Nonpoor	13,316	14,044	14,635	57.5	60.0	61.9
	\$1 to 24	4,121	5,401	5,825	17.8	23.1	24.6
	\$1-4	1,267	1,453	1,604	5.5	6.2	6.8
	\$5-9	1,023	1,062	1,531	4.4	4.5	6.5
	\$10-14	979	974	1,322	4.2	4.2	5.6
	\$15-19	852	1,127	703	3.7	4.8	3.0
	\$20-24	1,005	785	665	4.3	3.4	2.8
	\$25 to 49	2,523	2,567	2,431	10.9	11.0	10.3
	\$50 to 74	1,016	1,204	1,136	4.4	5.1	4.8
	\$75 to 149	1,645	1,516	1,709	7.1	6.5	7.2
	\$150-499	2,174	2,327	2,479	9.4	9.9	10.5
	\$500+	832	1,029	1,055	3.6	4.4	4.5

Table 4. (Continued) Absolute Value Difference Between Administrative and SIPP Monthly Social Security Income for SIPP Beneficiaries 65 Years and Older with \$1 or More Difference: March 1997, 1998, and 1999

Income for All Beneficiaries: Total and Split with \$1 or More Difference: March 1997, 1998, and 1999							
		March 1997	March 1998	March 1999	March 1997	March 1998	March 1999
		Numbers			Percent		
Total (Number in thousands)		25,788	25,900	26,221	100.0	100.0	100.0
Nonzero difference {SSA> or < SIPP}							
	Poor	1,611	1,612	1,705	60.8	63.4	66.5
\$1 to 24		623	588	646	23.5	23.1	25.2
	\$1-4	185	181	241	7.0	7.1	9.4
	\$5-9	125	131	209	4.7	5.1	8.2
	\$10-14	149	131	64	5.6	5.1	2.5
	\$15-19	128	79	76	4.8	3.1	3.0
	\$20-24	36	66	56	1.4	2.6	2.2
\$25 to 49		223	194	232	8.4	7.6	9.0
\$50 to 74		93	106	99	3.5	4.2	3.9
\$75 to 149		168	142	163	6.3	5.6	6.4
\$150-499		347	347	331	13.1	13.6	12.9
\$500+		157	235	234	5.9	9.2	9.1
Note: Dollar amounts in current (non-inflation-adjusted) dollars.							
Both SIPP and SSA social security amounts exclude Medicare Part B insurance.							
Universe is persons 65+ with positive social security benefit amounts in SIPP or SSA.							
Source: Survey of Income and Program Participation (SIPP) 1996 panel wave files and linked data from the Social Security Administration							

Nominal imprecision combined with SMI reporting error. Another group of beneficiaries to investigate were those who had a SSA-SIPP's social security benefit differential that corresponded to the level of SMI premium amount of approximately \$45 (\$45 to \$55). Roughly 5 percent of all elderly beneficiaries had differentials of this magnitude. In roughly half these cases the SIPP estimate was greater than the SSA amount. It could be that some of these beneficiaries had an approximate idea of their social security check amount and/or an approximate idea of the SMI amount. The summation of these two amounts, albeit an erroneous summation, would generate an amount such as they reported in SIPP. However, when checking to see if any SIPP respondents incorrectly reported a social security dollar amount that equaled the sum of the SSA benefit check plus the [most common] SMI amount, none was found. More investigations into these cases needs to be done to substantiate such claims of multiple sources of error, and therefore, they are excluded from any total error presented.

Error from Rounding

One segment of SIPP beneficiaries appeared to have rounded their SIPP social security amounts to the nearest \$10, \$100 or \$1000. Approximately 3.4 million appeared to have rounded in an easily observable way; that is, they rounded to the nearest \$10, \$100 or \$1000 (2.1 million) or rounded to the other, neighboring \$10, \$100 or \$1000 (1.3 million).¹¹ Interestingly, those who rounded to neighboring units and not the nearest units tended to round the SIPP estimate downward (1.1 million) (Table 5)

One might believe that those 2.1 million persons who rounded in the standard way properly excluded their SMI premium from the SIPP amount. It's also feasible to think that those who essentially "forced rounding downward" or "upward" (rounding to a neighboring \$10, \$100, or \$1000) also excluded the SMI amount in the social security amount they reported in SIPP.

Correcting for Errors and Elderly Poverty Rates

To correct for the errors, the SIPP social security amount was replaced with the SSA's direct pay which is the amount people get in their social security check –with any SMI premium deducted. The substitution was done for all family members, not just the elderly family members, since some of the elderly's family members also may be social security beneficiaries whose benefit amount may have been in error. Then the total family income was recalculated and their poverty rate was re-evaluated.¹¹ The poverty rate of the SIPP elderly social security beneficiaries in March 1999 declined from 9.8 percent to 8.5 percent after adjusting for errors in SIPP reporting. (Table 1)

Adding SMI Premium, Correcting for Errors and Poverty Rates

The method for adding SMI amounts back to the benefit check amount was to replace the SIPP social security beneficiary's dollar amount with the SSA direct pay amount and add to that the SSA's SMI premium amount. After this adjustment was made, the poverty rate of the elderly declined from 9.8 percent to 7.5 percent (Table 1)

¹¹ Following the Office of Management and Budget's (OMB) Statistical Policy Directive 14, the Census Bureau uses a set of annual money thresholds that vary by family size and composition. For SIPP, these thresholds were divided by 12 to arrive at monthly poverty threshold.

Conclusion

By correcting for the errors in the SIPP social security benefit check amounts, the poverty estimate for the elderly was significantly lowered. In addition, by adding the SMI income back to the check amount, the poverty rate also declined – yielding a combined improvement of 2.3 percentage points (9.8 percent compared with 7.5 percent). This finding calls for expansive and ongoing research with SIPP data; it has long been argued that excluding SMI income from SIPP's collection means that SIPP total money income may be underestimated and affect such key socioeconomic indicators such as the poverty status of the elderly. It would be important to search for any patterns of reporting. In addition, with SIPP social security income questions having been modified since the 1996 panel, it would be helpful to check the quality of respondents' answers in these instances. Another reason such research may become a greater topic of interest is that additional medicare coverage options are being extended to social security beneficiaries which will generally entail higher premium deductions.

It should be noted again that since not all SIPP elderly person records were matched to administrative data, those not matched may have some different characteristics than those matched. One example is that the poverty rate for elderly SIPP beneficiaries whose data were matched was 10.1 percent in March 1999, compared with a poverty rate of 14.7 percent for those whose data were not matched. (Table 1) If one had matched data for the latter group, then improvement results may or may not have been at the above-stated levels, but improvement would have taken place nonetheless. Potential research areas include understanding this non-matched universe better. In addition, efforts are underway at the Census Bureau to improve the matching process by augmenting current practices with new probabilistic methods. The Bureau is also exploring the feasibility of making use of administrative data in the editing process of social security income.

Interestingly, SMI premiums appeared to be properly excluded from roughly three-fourths of SIPP beneficiaries' social security amounts in the 1996 panel. That is to say, many SIPP beneficiaries excluded their SMI amount, as directed. This is definitely true for the 37.7 percent whose SIPP amount exactly matched the SSA amount. Yet, from the above discussion, it appears that this may also be true for the 35.2 percent possibly demonstrating low-level precision errors, rounding errors or COLA errors. No clear understanding may be given to the remaining 27.1 percent of elderly social security beneficiaries who exhibited some other, currently undetermined error. (Table)

These findings also give some weight to the need to look at other income sources collected by SIPP and other surveys for the existence and impact of specific types of nonsampling errors as preliminarily explored here. If methods such as using administrative records are not readily available to make corrections, then it may be important to note how correcting only some of the errors may tend to result in biased estimates.

	Number	Percent
Number (In thousands)	26,221	100.0
Specified type of error		
Total with one or more of specified error types	9,226	35.2
Cost of living adjustment not made in SIPP	1,410	5.4
Precision		
SIPP Less than \$25 different from SSA	6,469	24.7
SIPP less than \$10 different from SSA	3,584	13.7
Less than 5 percent different from SSA	7,640	29.1
Rounding	3,406	13.0
Rounding to nearest 10's 100's or 1000's	2,141	8.2
Rounding to other neighboring 10's 100's or 1000's	1,265	4.8
SIPP amount biased downward below SSA amount	1,054	4.0
SIPP amount biased upward above SSA amount	211	0.8
Unspecified type of error /1	7,113	27.1
No error found	9,882	37.7
/1 Error not identified in this paper.		
Note: If one sums together the specified types of error above, it will be greater than the total with one or more specified types shown in the table totals. For example, an erroneous SIPP social security income amount may demonstrate both imprecision error and rounding error, based on the definitions in this paper.		
Universe is persons 65+ with positive social security benefits in SIPP and/or SSA.		
Source: Survey of Income and Program Participation (SIPP) 1996 panel wave files and linked data from the Social Security Administration (SSA)		

References

Huynh, Minh, Kalman Rupp, and James Sears (2001). "The Assessment of Survey of Income and Program Participation (SIPP) Benefit Data Using Longitudinal Administrative Records," Proceeding of the Federal Committee on Statistical Methodology (FCSM) Research Conference, November 15, 2001, Key Bridge Marriott Hotel, Arlington, VA.

Vaughan, Denton (2003). Notes and discussions.