THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Two Notes on Sampling Variance Estimates from the 1984 SIPP Public-Use Files

No. 79

Barry V. Bye and Salvatore J. Gallicchio
Social Security Administration

April 1989

U.S. Department of Commerce U.S. CENSUS BUREAU

Survey of Income and Program Participation

Two Notes on Sampling Variance
Estimates from the 1984 SIPP
Public-Use Files

8902 79

Barry V. Bye and Salvatore J. Gallicchio
Social Security Administration

SIPP Working Paper #79

April 1989

The views expressed are the authors' and do not necessarily reflect those of the Census Bureau.

Preface

This working paper provides two contributions by Barry Bye and Sal Gallicchio of the Social Security Administration related to the estimation of variances from the SIPP public-use files. The 1984 public-use data files of the Survey of Income and Program Participation provide pseudo stratum and pseudo primary sampling unit codes that permit direct estimates of sampling errors. The first note is a reprint of an October 1988 Social Security Bulletin article describing a methodology for calculating sampling errors directly from the SIPP public-use file. The authors applied this method to the calculation of variances for persons participating in programs administered by the Social Security Administration, and empirically show an apparent sensitivity of generalized variances (as found in the SIPP Users' Guide and Technical Documentation) to curve fitting procedures.

The second note in this working paper reports the results of comparisons of direct variance estimates from the public-use file with variance estimates based on the original sample design (computed by Census Bureau staff). The authors conclude that the variance estimates are very much alike, suggesting some validity for the direct variance estimates using the pseudo design codes.

TABLE OF CONTENTS

Two Notes on Sampling Variance Estimates from the 1984 SIPP Public-Use Files

by Barry V. Bye and Salvatore J. Gallicchio*

The Census Bureau's Survey of Income and Program Participation (SIPP) provides data that can be used to study the characteristics of Old-Age, Survivors, and Disability Insurance (OASDI) and Supplemental Security Income (SSI) program participants. It is important that estimates of sampling errors accompany such studies because the estimates may have large sampling errors due to the small number of sample cases available for specific analyses. The generalized sampling variances provided by the Census Bureau did not identify separately either program's participants and, therefore, do not pertain directly to analyses of these groups. This article describes an approach to the direct computation of sampling variances for OASDI and SSI program participants. The approach uses the pseudo stratum and half-sample codes available in SIPP public use data files. A table of generalized standard errors is constructed for participants of both programs aged 18 or older. Generalized standard errors could not be computed for child beneficiaries under age 18 because of a wide variation of design effects across subpopulation estimates.

The Survey of Income and Program Participation (SIPP) provides data that can be used to study the socioeconomic characteristics of persons participating in programs administered by the Social Security Administration (SSA): Old-Age, Survivors, and Disability Insurance (OASDI) and Supplemental Security Income (SSI).[1] Currently, data from the initial 1984 SIPP panel are available. The 1984 panel consists of approximately 20,000 households comprising about 54,000 individuals. Through a special algorithm developed by SSA, about 8,000 of these individuals have been identified as OASDI and SSI program participants.[2] Included among them are about 4,600 retired-worker beneficiaries, about 600 disabled-worker beneficiaries, and 700 aged, blind, or disabled SSI recipients. The remaining participants are survivor, spouse, or child beneficiaries.

To provide summary SIPP data on SSA program participants to the public, a special set of tables was introduced in the Annual Statistical Supplement to the Social Security Bulletin for 1987.[3] The tables pertain to the civilian noninstitutionalized population receiving OASDI and SSI payments. They focus on three major themes: the composition and level of income of persons receiving different types of OASDI benefits, the general characteristics of persons aged 18-64 receiving OASDI or SSI payments based on disability, and similar information about SSI recipients aged 18 or older. The unit of analysis in these tables is the individual recipient.

Many of the distributions and income levels shown in the Supplement tables are based on a relatively small

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

[1] General information on the SIPP can be found in Dawn Nelson, David McMillen, and Daniel Kasprzyk, *An Overview of the Survey of Income and Program Participation* (SIPP Working Paper Series, No. 8401, update 1), Bureau of the Census, Department of Commerce, 1985.

[2] Denton R. Vaughan, *A Survey-Based Type of Benefit Code for the Social Security Program* (ORS Working Paper Series), Office of Research and Statistics, Social Security Administration (forthcoming).

[3] *Annual Statistical Supplement to the Social Security Bulletin, 1987*, Office of Research and Statistics, Social Security Administration, 1987, tables 15-22.

number of sample cases. Summary statistics generated from small numbers of cases can be imprecise due to large sampling errors (variances) and often suggest differences between subpopulations when no real differences exist. It is important, therefore, that estimates of sampling errors be provided along with the estimates of direct interest.

The Bureau of the Census has provided generalized variance curves for a number of quantities from the 1984 SIPP panel.[4] These curves do not identify OASDI or SSI recipients separately; therefore, the curves do not pertain directly to SSA program participants. Fortunately, provisions were made for the direct calculation of sampling variances of SIPP estimates using special codes available in the SIPP public use data files. These codes allocate the SIPP sample cases to a set of pseudo strata and pseudo primary sampling units. The codes permit direct estimates of sampling variances to be obtained by a number of methods.

The results of direct sampling variance computations for SSA program participants are presented in this article. The approach used to estimate the variances was the method of balanced half-sample replication.[5] The appendix at the end of the article includes the detailed specifications for estimating sampling variances from the SIPP using the same techniques that were used for the computations presented in this article. The results of the calculations also are provided in sufficient detail to be used as a benchmark.

Sampling variances were computed for more than 300 population estimates, cross-classifying the recipients by sex, age, marital status, and type of beneficiary. A curve was fit to the estimated variances and was used to produce tables of generalized standard errors. The tables of generalized standard errors can be applied directly to the data presented in the Supplement for program participants aged 18 or older and also can be used with other analyses from wave 1 of the 1984 SIPP panel that pertain to SSA program participation of adults. A separate analysis for child beneficiaries under age 18 showed that estimated standard errors were strongly associated with family size. As a result, tables of generalized standard errors that would be applicable to a variety of estimates for this subpopulation could not be developed.

The generalized variance curve presented in this article yields variance estimates that are markedly different from those generated by curves from the Census Bureau. In part, the difference may be due to the fact that variances of individual items estimated from the pseudo sample design may differ from those estimated directly from the original design. However, a part of the difference appears to be due to differences in the fit of the curves employed by the Census Bureau and by SSA staff, even though the functional form was the same. The SSA results appear to be more appropriate for variance estimates of OASDI and SSI program participants.

Sampling variances were also computed for some of the median income amounts shown in the Supplement. The variances and estimated sampling covariances between the medians were used to test hypotheses about differences in the size of the estimated median income amounts among various subpopulations.

[4] Survey of Income and Program Participation, User's Guide, Bureau of the Census, Department of Commerce, July 1987, pages 7-1 through 7-27.

[5] Kirk Wolter, Introduction to Variance Estimation, Springer-Verlag, New York, 1985.

Methodology

Balanced Half-Sample Replication

The method of balanced half-sample replication is an approach to the estimation of sampling variances for complex sample designs that can be implemented easily and has been applied to a wide variety of statistical estimates. This method presupposes that the primary sampling units for the population have been assigned to one of L strata, and two of the units are selected with replacement from each stratum with probability proportionate to size. Half-sample replicates of this design can be formed by selecting one of the two units from each stratum. For a sample design with L strata, there are 2^L such half samples. If an estimate of the statistic of interest is made in each half sample and in the full sample, then the average squared difference between half-sample and full-sample estimates from any subset of half samples provides an estimate of the sampling variance of the statistic. The estimate of the sampling variance is most precise when all 2^L half samples are employed.

When L is large, one would like to use only a part of the 2^L half samples to estimate the sampling variances without loss of precision. It turns out that special sets of half samples, called balanced, orthogonal sets, are particularly good candidates. Estimates of sampling variances from these special sets are algebraically equivalent to those obtained using all half samples. Also, when the full-sample estimate is a linear function of the half-sample estimates, the average estimate over the balanced, orthogonal set will be equal to the full-sample estimate. The minimum number of half samples required for a fully balanced orthogonal set is the smallest multiple of 4 which is greater than the number of strata in the sample design. For designs with many strata, this number will be much smaller than the total number of possible half samples. Descriptions of balanced, orthogonal sets for many designs are provided in the literature.[6]

Once a set of half samples has been identified, estimated sampling variances are particularly easy to compute. Let $\theta_\alpha(\alpha=1, \dots, K)$ denote the estimator of the population parameter of interest computed from the α th half sample, and let θ be the corresponding estimate from the full sample. An estimator of the sampling variance of θ, $V(\theta)$, based on K half samples is given by

$$V(\theta) = \sum_{\alpha=1}^{K} (\theta_\alpha - \theta)^2/K \qquad (1)$$

When θ is a linear function of the θ_α, so that

$$\theta = \bar{\theta} = \sum_{\alpha=1}^{K} \theta_\alpha/K \,,$$

then (1) provides an unbiased estimate of the variance of θ. When θ is not linear in θ_α (for example, θ is a ratio, a median, a correlation coefficient), then $\theta \neq \bar{\theta}$ and the expected value of $V(\theta)$ differs from the variance of θ by an amount often well approximated by $[E(\bar{\theta} - \theta)]^2$. Thus if $\bar{\theta}$ is close to θ, equation (1) will provide a good approximation of the sampling variance when θ is not linear.[7]

Variance Curve

A two-parameter curve was fit to the variance estimates obtained by the replication method. The curve specified the relative variance (Rv), the variance divided by the square of the estimate, as a function of the estimate.

$$Rv(x) = a + b/x \qquad (2)$$

where

a and b are coefficients to be estimated,
x is the estimated population total, and
Rv(x) is the estimated relative variance of x— that is,

$$Rv(x) = V(x)/x^2 \,.$$

[6]R. L. Plackett and J. P. Burman, "The Design of Optimum Multifactor Experiments," Biometrika, 33(1946), pages 305 and 325.
[7]Wolter (1985), op. cit., references a number of empirical investigations supporting the use of equation (1).

This functional form has provided a fairly good representation of the relationship between Rv(x) and x in other surveys. Its use is motivated by the following considerations.[8]

The design effect (Deff) for a particular estimate, x, from a complex sample design is defined as the ratio of the sampling variance of x under the design to the sampling variance that would have been obtained from a simple random sample of equal size. For a sample of size n from a population of size N, the simple random sampling variance of an estimated total, x is given by

$$var(x) = var(pN) = N^2PQ/n$$

where

P = X/N, is the true population proportion,
X is the population total estimated by x,
Q = 1-P, and
p is the sample estimate of P.

The variance of x from a complex design of the same size can be expressed as

$$var_c(x) = Deff\,(var(x)) = Deff\,(N^2PQ/n).$$

The relative variance of x is given by

$$Rv(x) = var_c(x)/X^2 = Deff\,(Q/Pn)$$
$$= -Deff/n + (N/n)Deff/X. \qquad (3)$$

Equation (3) has the same form as equation (2) where a = -Deff/n and b = (N/n)Deff. If it is reasonable to assume that a constant design effect exists for a particular set of estimates, then the estimated relative variances for those items may be accurately represented by a two-term curve of the form in (2) from which generalized variances can be computed.

The method used to estimate the coefficients in (2) was an iterative procedure that minimized the function

$$\sum_{i=1}^{I} \left[\frac{Rv_i - \hat{Rv}_i}{\hat{Rv}_i^*}\right]^2$$

where

Rv_i is the computed relative variance for the ith item;
$\hat{Rv}_i$ is the estimated relative variance for the ith item from the curve;

[8]See, for example, The Current Population Survey: Design and Methodology (Technical Paper 40), Bureau of the Census, Department of Commerce, January 1978.

$\hat{Rv}_i^*$ is a weight for the ith item. It is set equal to the computed relative variance, Rv_i, in the first iteration; for all subsequent iterations it is set equal to the estimated relative variance, $\hat{Rv}_i$, from the previous iteration.

I is the number of items to be fit.

This estimation approach gives greater weight to items with smaller estimated relative variances (and, thus, generally larger estimated totals) and has been found to work well in other surveys.[9]

Generalized Variances for Counts and Proportions

Having estimated values for the coefficients in equation (2), the relative variance for a specific estimated total, x_0, can be obtained by substituting x_0 into that equation. The variance of the estimated total can be obtained by multiplying the relative variance by the square of the estimate.

$$\hat{V}(x_0) = \hat{Rv}(x_0)x_0^2$$

$$= ax_0^2 + bx_0 \qquad (4)$$

Equation (4) can also be used to produce generalized estimates of variances of proportions. A proportion is the ratio of two estimated totals, $p = x/y$, where the cases counted in the numerator are a subset of the cases counted in the denominator. In large samples, the relative variance of this type of ratio can be approximated by the following formula:

$$Rv(p) = Rv(x/y) = Rv(x) - Rv(y)$$

or

$$V(p) = V(x/y) = (x/y)^2\,[Rv(x) - Rv(y)] \qquad (5)$$

Substitution of estimates from (2) into (5) provides generalized variance estimates for proportions.

$$\hat{V}(p) = p^2[b(1/x - 1/y)] = (b/y)\,(p)\,(1 - p)\,. \qquad (6)$$

Tables of generalized standard errors for estimated totals are often produced from equation (4) by computing and displaying the square root of the estimated variances for a set of predetermined values of x. Similarly, a table of standard errors for estimated proportions can be computed from (6). This table will be two dimensional with the size of the base of the percent on one dimension and the estimated proportion on the other.

Variances of Medians

The balanced half-sample replication approach was used to estimate standard errors for the estimated medians in table 17 of the 1987 Supplement. That table presents median OASDI income, median total income, and the median of the ratio of OASDI income to total income for several beneficiary groups, cross-classified by a number of factors.

In this article, the medians were estimated from distributions of the variables of interest using the following formula:[10]

$$M = L_j + \left[\frac{S_{50} - S_j}{N_j}\right] W_j$$

where

- j indexes the interval containing the 50th percentile;
- L_j is the lower limit of the jth interval;
- S_{50} is the estimated population at the 50th percentile;
- S_j is the estimated population with values below the jth interval;
- N_j is the estimated population in the jth interval; and
- W_j is the width of the jth interval.

An interval width of $25 was used for the OASDI income distribution. Intervals of $50 or $100 were employed for the total income distribution, the latter used to capture the larger monthly benefit amounts. An interval of .05 was used for the income ratio.

The sampling variance of M was obtained by estimating M in each half sample and then applying

---

[9] There is no specific justification for this weighted least squares approach other than the usefulness of its results. Ordinary least squares estimates, minimizing

$$\sum_{i=1}^{I} (Rv_i - \hat{Rv}_i)^2\,,$$

have been found to give too much weight to small estimates, x, with characteristically large estimated relative variances. Nonlinear least squares estimates, minimizing

$$\sum_{i=1}^{N} \left[\frac{Rv_i - \hat{Rv}_i}{\hat{Rv}_i}\right]^2,$$

appear to give too much weight to observations with large estimated totals.

[10] The estimated medians shown in the Supplement were computed by the TPL tabulation program on an IBM system. The medians reported here were computed by the PASS tabulation program on a UNIVAC system and they sometimes differ from the Supplement estimates by small amounts.

equation (1). This approach was repeated for each of the three median amounts and for each subpopulation.

Statistical Tests for Differences of Medians

Statistical tests were made on the variation in medians across the categories of a particular variable (sex, age, and size of family, for example) within a particular beneficiary group. The test approach follows that developed by Grizzle, Starmer, and Koch.[11] Let $M_1, M_2, \ldots, M_k$ be a set of estimated medians for k categories of the variable. Then a χ^2— type test statistic for the hypothesis $H_0 : M_1 = M_2 = \ldots = M_k$ can be constructed under the assumptions that the M have, jointly, a multivariate normal distribution and that a consistent estimate of the sampling covariance matrix is available.[12]

The sampling covariance matrix is obtained through the balanced half-sample method by a computation similar to that of equation (1). The (i,j)th element of the matrix is given by

$$\sum_{\alpha=1}^{K} [M_\alpha^{(i)} - M^{(i)}]\,[M_\alpha^{(j)} - M^{(j)}]/K .$$

where

$M^{(r)}$ is the estimate of the median for the rth category from the entire population,

$M_\alpha^{(r)}$ is the estimate of the median for the rth category from the α th half sample, and

K is the number of half samples.

Among retired-worker beneficiaries, in two cases, the set of categories consists of a cross-classification of two factors: sex by age and sex by marital status. In these cases, a sex effect, an age (or marital status) effect and a combined effect were tested. For disabled-worker beneficiaries, the type-of-family categories refer to both marital status and presence of minor children. In this case, the medians for married versus not married and the medians for married with minor children versus married with no minor children were tested.

[11]J. R. Grizzle, C. F. Starmer, and G. C. Koch, "Analysis of Categorical Data by Linear Models," Biometrics, September 1969, pages 489-504.

[12]The asymptotic normality of the estimated medians follows from the asymptotic normality of the estimated ratios $(S_{50}/N_j, S_i/N_j)$ of which the median is a linear function. The covariance matrix computed by half-sample replication on the pseudo design is not a consistent estimate. Still, it is believed that the GSK test statistics provide useful information about the real spread in the medians, even if the true significance levels are not known.

Results

Participants Aged 18 or Older

Appendix table I presents the population estimates, standard errors, and relative variances for each of the items described above. There were 326 subpopulation estimates based on more than 1 sample case. The estimates ranged from a low of about 7,000 based on 2 sample cases to a high of 38 million based on 7,943 sample cases that represent the entire OASDI and SSI recipient population.[13] The variance curve that was dervied from the items has coefficients[14]

$$a = .0007$$
$$b = 5217.$$

Tables of generalized standard errors based on this curve follow.[15] For the estimated totals of a specific size, table 1 gives one standard error of the estimate. Table 2 gives one standard error for estimated proportions with bases of various sizes.

Participants Under Age 18

When constructing estimates of family characteristics for children, one would expect large design effects in the estimated sampling errors. All children will tend to report (or have coded for them) the same family data, thus reducing the effective number of independent observations by the average number of children per family. Because OASDI benefits awarded to minor children tend to be divided among all the children in a beneficiary family, the strong clustering effects that one finds for child-related estimates are expected to appear for beneficiary children as well.

To investigate the sampling variances for children, a set of estimates was constructed by cross-classifying

[13]A sampling variance cannot be estimated for totals based on 1 sample case. Algebraically, the balanced half-sample estimator yields a perfect 1.0 for the estimated relative variance. Thirty-nine of these cells are shown in appendix table I.

[14]The estimated constant, a, is positive. Although the rationale presented suggests that a should be negative, the algorithm used to estimate the parameters does not impose this constraint. The estimated design effect from the b coefficient is

$$Deff = b(n/N) = (5217)\,(7943/34160810) = 1.2.$$

Values for n and N are obtained from the first item in the variance table in the appendix.

[15]Variance curves were also estimated for sets of items for several subpopulations of the total beneficiary population: disabled workers, persons aged 65 or older, and persons receiving SSI payments. Generally, the sizes of standard errors for similar size cells across these groups did not differ. A curve was also estimated for the group aged 18 or older, using items derived from cross-classifying age, family size, and family income. Again, no substantial differences were seen in estimated a and b parameters.

family size, family income, sex, and race. As expected, a variance curve fit to all of the items exhibited a systematic lack of fit, overestimating the computed variances for smaller families and underestimating the variances for larger families. Fitting separate curves by family size resulted in the following set of a and b parameters:

	Parameter	
Family size	a	b
1-3	.0034	4922.
4	.0127	5849.
5 or more	.0199	8733.

The increasing values of both the a and b parameters indicate that substantial increases in sampling variances are to be expected, for an estimate of fixed size, as family size increases.

Table 1.—Standard errors for estimated population totals

Estimate	Standard error
25,000	11,436
50,000	16,202
75,000	19,878
100,000	22,994
250,000	36,738
500,000	52,842
750,000	65,786
1,000,000	77,176
2,500,000	132,954
5,000,000	211,284
7,500,000	284,417
10,000,000	355,574
25,000,000	771,177
50,000,000	1,455,403

These results imply that the sampling variance for an estimated subpopulation of child beneficiaries under age 18 will depend largely on the family size composition of the subpopulation. A set of child-beneficiary estimates would not be likely to exhibit a constant design effect; and therefore, it is unlikely that a two-term curve of the kind described above would provide a good approximation to the estimated sampling variances for the set. Accordingly, no generalized variances for child beneficiaries are presented. There appears to be no substitute for direct variance calculations in this case.

Comparison with Census Generalized Variances

The SIPP User's Guide [redacted] Census curve 1, [redacted]

Table 3 shows estimated standard errors from the SSA curve and Census curve 1 for a range of estimates.[17] For estimates less than 10 million, the Census estimates are 1.20 to 1.75 times larger than those from the SSA curve. Some of this difference could be due to differences in computational schemes for the direct

[16]SIPP User's Guide, op. cit., page 7-5.

[17]The parameters from Census curve 1 are:

a = -.0000942, and b = 16059.

Table 2.—Standard errors for estimated percents

Base of percent	Percent											
	1 or 99	2 or 98	5 or 95	8 or 92	10 or 90	15 or 85	20 or 80	25 or 75	30 or 70	35 or 65	40 or 60	50
25,000	4.54	6.39	9.95	12.39	13.70	16.31	18.27	19.77	20.93	21.78	22.37	22.83
50,000	3.21	4.52	7.04	8.76	9.69	11.53	12.92	13.98	14.80	15.40	15.02	16.14
75,000	2.62	3.69	5.75	7.15	7.91	9.41	10.55	11.42	12.08	12.58	12.92	13.18
100,000	2.27	3.20	4.98	6.19	6.85	8.15	9.13	9.89	10.46	10.89	11.19	11.42
250,000	1.44	2.02	3.15	3.92	4.33	5.16	5.78	6.25	6.62	6.89	7.07	7.22
500,000	1.02	1.43	2.23	2.77	3.06	3.65	4.00	4.42	4.68	4.87	5.00	5.11
750,000	.83	1.17	1.82	2.26	2.50	2.98	3.33	3.61	3.82	3.98	4.08	4.17
1,000,000	.72	1.01	1.57	1.96	2.17	2.58	2.89	3.13	3.31	3.44	3.54	3.61
2,500,000	.45	.64	1.00	1.24	1.37	1.63	1.83	1.98	2.09	2.18	2.24	2.20
5,000,000	.32	.45	.70	.88	.97	1.15	1.29	1.40	1.48	1.54	1.58	1.61
7,500,000	.26	.37	.57	.72	.79	.94	1.05	1.14	1.21	1.26	1.29	1.32
10,000,000	.23	.32	.50	.62	.68	.82	.91	.99	1.05	1.09	1.12	1.14
25,000,000	.14	.20	.31	.39	.43	.52	.58	.63	.66	.69	.71	.72
50,000,000	.10	.14	.22	.28	.31	.36	.41	.44	.47	.49	.50	.51

Table 3.—Comparison of generalized standard errors for estimated totals

Estimate	SSA	Census curve 1	Percent
25,000	11440	20035	175.1
50,000	16206	28332	174.8
75,000	19882	34697	174.5
100,000	22997	40062	174.2
250,000	36731	63316	172.4
50,0000	52805	89476	169.4
750,000	65708	109505	166.7
100,000	77051	126352	164.0
250,000	132446	198894	150.2
500,000	209962	279177	133.0
750,000	282181	339328	120.3
10,000,000	352375	388806	110.3
25,000,000	761853	585320	76.8

variance estimates on which the curves are based. Both the variance estimators and the assumed sample design are different.[18]

Much of the difference in the curves, however, appears to be attributable to differences in curve-fitting strategies. The Census curve is based on 36 estimated totals for persons aged 16 or older involving receipt of cash and noncash benefits and labor-force activity. Thirteen of the 36 items are estimates of the Hispanic population with selected characteristics. Unpublished Census Bureau data suggest that variances from curve 1 for population totals of less than 500,000 are substantially overestimated.[19] This is not surprising because only several observations are in this range among the 36 items and they are given little weight by the kind of curve-fitting algorithm described above.[20] As indicated in the appendix, the set of items from which the SSA curve was derived contains a large number of small estimates. The SSA curve appears to fit the observations well for small estimated totals.

The reasons for differences between Census Bureau and SSA curves for larger estimates are more difficult to discern. There is some indication that the design effects for the Hispanic population estimates are larger than those for the corresponding estimates for all races combined, raising the overall level of the Census curve. It is also possible that the design effects for adult OASDI and SSI program participants are generally smaller than the effects for the Census items. Less clustering may occur among OASDI and SSI adult recipients in families and households, compared with recipients in other transfer programs. The small number of items on which the Census curve is based makes a more detailed analysis difficult. At this point, the SSA curve appears to be much preferred for OASDI and SSI program participation estimates.

Medians

The standard errors for the medians in table 17 of the Annual Statistical Supplement are shown in table 4. With the exception of child beneficiaries, the variances of the estimated medians appear to be quite small. The sizes of the estimated standard errors rarely exceed 10 percent of the corresponding medians and are often well under 5 percent. The median income amounts for families of child beneficiaries show larger standard errors than, for example, similar estimates for families of disabled-worker beneficiaries even when the unweighted case counts are about the same. The larger estimated standard errors are probably the result of the clustering effects for child beneficiaries discussed above.

The generally small standard errors are also reflected in the test statistics for the hypotheses concerning differences of medians. For each set of categories and each type of median, the differences between medians across categories were statistically significant at the .05 level in most cases. When contrasts were significant, the significance levels tended to be much smaller than .05, usually less than .0001.

The contrasts that were not significant at the .05 level are described at the end of table 4. The table identifies the specific comparisons and provides the value of the test statistic, the degrees of freedom, and the p-value. The following examples demonstrate how the test results can be interpreted.

The statistical tests indicated no two-way interaction existed between sex and age regarding the ratio of OASDI benefits to total income for retired-worker beneficiaries. Differences in median ratios between age groups tended to be about the same for both men and women. The differences between median ratios for men by age group are 13, 9, and 0. The corresponding differences for women are very similar (12, 7, and 2). The statistical tests did show significant sex differences and significant age differences. The pattern of median ratios, therefore, can be described by adding sex and age effects without the need to adjust for particular sex-age combinations.

[18] Census estimates were computed by the half-sample replication method using a set of 50 half samples that was not fully balanced. The appendix provides a brief description of the procedures used to create the pseudo design codes.

[19] For a description of the items, see "Memorandum for Documentation from Karen E. King, Subject: SIPP Variances: Items by Generalized Variance Parameter," Bureau of the Census, Department of Commerce, June 19, 1985. The Census direct variance estimates are unpublished and were made available by the Statistical Methods Division, Bureau of the Census.

[20] The Census Bureau curve-fitting algorithm differed from that described above in that the relative variance for the overall population total, T, was constrained to be zero. Thus, $a + b/T = 0$ or $a = -b/T$, and b is estimated from a one parameter model $V(x) = b(1/x - 1/T)$. This approach is reasonable because the case weights are adjusted to achieve certain population totals. However, imposing this constraint may also contribute to the overestimate of the variance for small population estimates.

Table 4.—Standard errors for table 17, Annual Statistical Supplement to the Social Security Bulletin, 1987

Characteristic	OASDI benefit		Total income		Ratio*	
	Median	Standard error	Median	Standard error	Median	Standard error
	Retired workers					
Total	577	10	1210	23	53	1
Men	633	10	1300	30	51	1
Women	515	7	1096	29	57	1
Sex and age of beneficiary:[1]						
Men—						
62-64	502	11	1442	54	34	2
65-69	672	18	1444	51	47	2
70-74	682	13	1282	40	56	2
75 or older	611	16	1137	35	56	1
Women—						
62-64	582	39	1481	76	41	2
65-69	569	19	1216	28	53	2
70-74	531	12	1072	42	60	2
75 or older	469	9	847	45	62	2
Sex and marital status:						
Men—						
Married	697	9	1417	26	50	1
Widowed	456	13	946	64	49	2
Divorced	451	33	759	93	64	4
Never married	476	34	893	79	56	3
Women—						
Married	763	8	1487	38	52	2
Widowed	437	6	760	28	61	2
Divorced	411	13	778	57	58	4
Never married	452	20	935	115	58	3
Size of family:						
1 person	419	6	629	19	65	1
2 persons	713	9	1351	28	54	1
3 persons or more	669	29	2261	74	30	1
Monthly family income:						
Less than $500	326	7	396	6	91	1
$500-$999	520	5	743	7	74	1
$1,000-$1,499	713	15	1225	7	57	1
$1,500-$1,999	718	15	1722	14	41	1
$2,000-$2,499	793	13	2203	13	35	1
$2,500-$2,999	710	41	2776	20	25	1
$3,000 or more	764	29	3891	83	17	1
Family source of income:						
Earnings[2]—						
Yes	572	15	1946	36	31	1
No	580	13	1015	29	63	1
Assets—						
Yes	622	9	1337	26	50	1
No	428	11	604	24	75	2
Means-tested cash benefits—						
Yes	335	16	594	56	58	1
No	600	9	1247	20	53	1
Other cash income—						
Yes	651	11	1461	23	46	
No	497	7	795	24	71	2
	Disabled workers					
Total	522	14	1162	47	49	2
Men[3]	566	12	1175	57	50	3
Women	419	26	1137	59	46	4
Age of beneficiary:[4]						
18-54	544	16	1240	83	45	4
55-64	501	18	1127	53	50	3

See footnotes at end of table.

Table 4.—Standard errors for table 17, Annual Statistical Supplement to the Social Security Bulletin, 1987—Continued

Characteristic	OASDI benefit		Total income		Ratio[a]	
	Median	Standard error	Median	Standard error	Median	Standard error
	Disabled workers—cont.					
Size of family:						
1 person	392	26	490	39	79	5
2 persons	547	21	1202	51	44	3
3 persons or more	597	25	1625	162	39	3
Type of family:[5]						
Married	578	15	1367	97	44	2
With minor children	713	48	1284	125	54	6
No minor children	547	17	1427	115	41	3
Unmarried	434	21	833	50	55	5
Monthly family income:						
Less than $1,000	437	19	620	42	80	3
$1,000-$1,999	616	20	1369	49	44	2
$2,000 or more	563	43	2664	113	18	1
Family source of income:						
Earnings[6]—						
Yes	516	17	1831	69	31	2
No	528	20	803	50	70	3
Assets—						
Yes	566	23	1512	90	41	2
No	483	16	822	53	63	4
Means-tested cash benefits—						
Yes	407	34	858	67	52	4
No	553	16	1266	65	47	3
Other cash income—						
Yes	594	20	1574	75	41	2
No	477	14	884	48	62	5
	Nondisabled widows					
Total	379	8	657	33	59	2
Age of beneficiary:[7]						
60-69	363	12	834	43	47	3
70 or older	386	9	579	25	68	3
Size of family:						
1 person	363	10	471	18	72	2
2 persons	458	19	1227	82	41	5
3 persons or more	373	15	2104	210	17	2
Monthly family income:						
Less than $1,000	361	9	478	10	79	2
$1,000-$1,999	443	21	1304	36	32	2
$2,000 or more	401	16	2939	84	13	1
Family source of income:						
Earnings[6]—						
Yes	368	10	1759	184	19	2
No	385	10	496	20	75	2
Assets—						
Yes	403	7	825	38	51	2
No	316	11	405	15	81	3
Means-tested cash benefits—						
Yes[8]	258	12	454	32	59	3
No	396	7	706	34	59	3
Other cash income—						
Yes	406	16	1033	69	39	2
No	369	8	525	21	72	2

See footnotes at end of table.

Table 4.—Standard errors for table 17, Annual Statistical Supplement to the Social Security Bulletin, 1987—Continued

Characteristic	OASDI benefit		Total income		Ratio*	
	Median	Standard error	Median	Standard error	Median	Standard error
	Minor children					
Total	604	41	1463	114	43	3
Size of family:[10]						
1 or 2 persons	392	61	981	132	43	11
3 persons	622	77	1437	155	50	7
4 persons	674	69	1578	252	46	10
5 persons	543	101	1800	198	30	5
6 persons or more	539	90	1345	213	45	5
Type of family:[11]						
With husband/wife head	601	42	1828	112	32	3
With single head	615	75	1181	70	49	5
Monthly family income:						
Less than $1,000	464	33	674	57	81	5
$1,000-$1,999	700	48	1449	79	46	3
$2,000 or more	675	89	2928	189	20	3
Family source of income:						
Earnings—						
Yes	519	34	1829	78	31	2
No	728	61	958	48	86	6
Assets[12]—						
Yes	655	53	1999	99	30	3
No	525	43	973	66	70	5
Means-tested cash benefits[13]—						
Yes	454	42	966	150	56	8
No	657	35	1713	133	39	3
Other cash income[14]—						
Yes	645	56	1911	66	34	3
No	541	50	1251	86	49	7

* OASDI divided by total; two decimals implied.

Finding	Chi**2	d.f.	p-value
[1]No two-way interaction in ratio	1.25	3	.74
[2]No difference in OASDI benefit level	.50	1	.70
[3]No difference in total income	.27	1	.60
No difference in ratio	.90	1	.34
[4]No difference in total income	1.56	1	.21
[5]No difference in total income for married with minor/with no minor	.69	1	.41
No difference in ratio for married with minor/with no minor	3.58	1	.06
[6]No difference in OASDI benefit level	.22	1	.64
[7]No difference in OASDI benefit level	2.60	1	.11
[8]No difference in OASDI benefit level	1.54	1	.22
[9]No difference in ratio	.02	1	.89
[10]No difference in ratio	7.26	4	.12
[11]No difference in OASDI benefit level	.02	1	.88
[12]No difference in OASDI benefit level	3.02	1	.08
[13]No difference in ratio	3.73	1	.05
[14]No difference in OASDI benefit level	1.56	1	.21

In contrast to the sex-age findings for retired workers, the sex by marital status tests showed that a two-way interaction was required to describe the patterns of median ratios. Again, differences were seen among the medians for each factor separately, but the pattern of marital status differences was not the same for men and women. Note, for example, that the difference in median ratios for married men and widowed men, -1, appears to be quite different from the difference between the medians of married and widowed women, +9. Among the other sequential contrasts differences were also evident. This pattern of values can not be explained by additive effects alone.

Sex and age contrasts for disabled-worker beneficiaries present situations in which a significant difference existed among median OASDI benefits but not among total incomes or ratios. This apparent inconsistency could be due to chance alone. However, there could be another explanation. The median ratio is not, algebraically, the same quantity as the ratio of the medians. It is possible that the ratios of the medians in the population are different, as suggested by the data presented here, but that the median ratios in the population are the same.

The remaining findings of differences in medians generally indicate that a contrast between one pair of medians was not significant. The one exception is the contrast of family size ratios for families with minor children. Because there were five family size categories, four contrasts were involved in the comparison.

Conclusion

This article described a methodology for calculating sampling errors directly from the SIPP public use file and applied this method to the calculation of variances for persons participating in SSA-administered programs. The methodology is presented in sufficient detail so that researchers can apply the same methods to their specific analyses. Since the replication variance estimation approach is not difficult to implement and facilitates a wide range of hypothesis testing techniques, it is recommended that direct variance calculations be used. This position is further supported by the apparent sensitivity of generalized variances to curve-fitting procedures. Estimating variances directly will also permit variances to be obtained from subsequent waves of the 1984 SIPP panel. Presumably, estimated standard errors will be higher for later waves of the panel due to the accumulated sample attrition at each wave.

For those who cannot compute variances directly, standard error tables have been provided for OASDI and SSI program participants aged 18 or older from wave 1 of the 1984 panel. The standard errors pertain directly to the SIPP tables in the Annual Statistical Supplement to the Social Security Bulletin for 1987. The standard error tables can also be used for other analyses of program participants from wave 1. Generalized standard errors for participants under age 18 could not be developed.

Several matters need further investigation to raise confidence in direct sampling error estimates from the public use files. A comparison of variance estimates from the pseudo design and from the actual sample design will show whether the pseudo design yields estimates that are, on average, smaller than those obtained when the original design is used. A comparison of the size of test statistics of the type that are used in this article also would be useful. These statistics require estimates of sampling variances and covariances, and it would be helpful to know if the pseudo design yields reasonable estimates of covariance as well as variance. Finally, little is known about the raw sample sizes required before normality is achieved in the sampling distribution of the various statistics presented. If for small samples the sampling distribution of counts, proportions, or medians is markedly different from normal, it might be misleading to form confidence intervals or to perform statistical tests assuming a normal distribution (that is, assuming symmetric intervals of 1 standard error about the estimate yields a 68-percent confidence interval, 2 standard errors provides a 95-percent confidence interval). The true confidence intervals may be larger or smaller than those of a normal distribution and may not be symmetric about the estimate. All of these matters are important if the Survey of Income and Program Participation is to be used for making inferences about the population under SSA-administered programs and not just for descriptive reporting.

Appendix: Detailed Sampling Variance Specifications

Assignment of Half-Sample Codes

Each person in the sample in the 1984 SIPP public use file had been assigned a pseudo-stratum code and a pseudo primary sampling unit (PSU) code within each pseudo stratum.[1] Generally, a self-representing (SR) PSU from the original design was associated with two non-self-representing (NSR) PSUs to form a pseudo stratum. Segments of the SR PSU were assigned to one of the two pseudo PSUs at random; each of the NSR PSUs was assigned, in its entirety, to one or the other of the pseudo units. In some cases, two SR PSUs or four NSR PSUs were grouped to form a pseudo stratum. The assignment resulted in the formation of 71 pseudo strata with 2 pseudo units in each stratum. The original PSU codes were withheld from the public use file to prevent access to small geographic areas where a risk of disclosure of individual identities might be possible.

For a design with 71 strata with two units each, the smallest number of half samples that can achieve full orthogonal balance is 72. The set of balanced half samples used in the variance computations is shown in chart I.[2] The array represents a string of 72 1s and 0s for each of the 71 pseudo strata. For a SIPP sample case in pseudo-stratum δ and pseudo-unit 1, the string in the δ th row of the array was attached to the record. For a SIPP sample case in pseudo-stratum δ and pseudo-unit 2, the complement (that is, 1s replaced by 0s, and vice versa) of the string in the δ th row of the array was attached. These strings effectively assign each SIPP case to 36 of the 72 half samples. A "1" in the a th position in the string indicates that the case is to be included in the a th half sample; a "0" means that the case is not to be included.

Item Specification for Generalized Variances

Replication variances were obtained for estimated population totals of OASDI and SSI recipients. Recipiency status was determined by the responses for September 1983. Estimated population totals were obtained in each half sample by multiplying the sum of the weights by 2. [3] The recipients were cross-classified by age, sex, marital status, and type of recipient (OASDI only, SSI only, and concurrent OASDI and SSI). This cross-classification yielded 326 distinct detailed and subtotal cells with more than one case.

The September 1983, OASDI and SSI recipient universe consists of those persons in the sample who meet the following test:[4]

[(IO1AMT-* >0 or IO3AMT-* >0)
or
(SOCSEC-* = 1 and AGE-* < 18)]
and
[FNLWGT-* >0]

where

IO1AMT-*	refers to the OASDI benefit amount;
IO3AMT-*	refers to the SSI amount;
SOCSEC-*	is the OASDI indicator;
AGE-*	is age in September 1983; and
FNLWGT-*	is the case weight.

Each variable is selected for September based on the rotation group of the sample case shown below:

Rotation group	Month
1............................	4
2............................	3
3............................	2
4............................	1

The cross-classifying variables (type of benefit, age, sex, and marital status) were constructed as follows:

Age (AGE-*):

Under 18	65-69
18-24	70-74
25-34	75-84
35-44	85 or older
55-64	

Sex:
Male, Female

Type of benefit:

OASDI only..........	(IO1AMT-* > 0 and IO3AMT-* =0) or (SOCSEC-* = 1 and AGE-* < 18)
SSI only...............	(IO1AMT-* = 0 and IO3AMT-*> 0)
OASDI and SSI........	(IO1AMT-*> 0 and IO3AMT-*> 0)

Marital status (MS-*):	Code
Married...............	Under 2
Widowed..............	3
Separated..............	4, 5
Never married..........	6 or over

Table I presents the estimated sampling variances for the 326 items described above.

[1]These fields are identified as H*-STRAT and H*-HSC in the public use file data dictionary. The codes for month 1 were used. The codes do not vary by month.

[2]The 72 order design in Plackett and Burman (1946), op.cit., was used. The array can be generated by shifting the first row one digit to the left for each subsequent row.

[3]This estimator does not fully replicate the original SIPP estimator in each half sample. The original SIPP estimator consisted of a number of multiplicative adjustments to the raw case weights. Similar adjustments should have been applied separately in each half sample to properly replicate the full sample estimator. The overall effect on the estimated variance of not having done this is unknown.

[4]All variables are referred to by their public use file variable names.

Table I.—Variance estimates for OASDI and SSI participants under SSA-administered programs

Age	Sex	Marital status[1]	Unweighted count	Estimate	Standard error	Relative variance
			All program participants			
Total	Total	Total	7943	34160810.	883445.	.0006688
Total	Total	NM	1147	4938770.	207858.	.0017713
Total	Total	S	497	2291038.	99936.	.0019027
Total	Total	W	2307	9917379.	305171.	.0009469
Total	Total	M	3992	17013620.	568181.	.0011153
			OASDI only			
Total	Total	Total	7242	31012390.	814853.	.0006904
Total	Total	NM	973	4148071.	191974.	.0021419
Total	Total	S	358	1634194.	91508.	.0031356
Total	Total	W	2078	8966302.	277238.	.0009560
Total	Total	M	3833	16263820.	556481.	.0011707
<18	Male	NM	252	1051521.	89736.	.0072828
<18	Female	NM	256	1064085.	87690.	.0067913
<18	Total	NM	508	2115606.	146801.	.0048149
18-24	Male	S	1	46.	5646.	1.0000000
18-24	Male	NM	30	139714.	27131.	.0377100
18-24	Male	Total	31	145360.	28694.	.0389663
18-24	Female	W	3	10502.	6079.	.3350419
18-24	Female	NM	26	112174.	19133.	.0290918
18-24	Female	Total	29	122676.	20793.	.0287286
18-24	Total	NM	56	251888.	34246.	.0184839
18-24	Total	Total	60	268036.	36677.	.0187243
25-34	Male	M	6	29086.	12232.	.1768577
25-34	Male	W	1	4053.	4053.	1.0000000
25-34	Male	S	3	31835.	24101.	.5731619
25-34	Male	NM	16	89563.	23121.	.0666412
25-34	Male	Total	26	154536.	33560.	.0471601
25-34	Female	M	10	47962.	16933.	.1246478
25-34	Female	W	16	71050.	16858.	.0562995
25-34	Female	S	1	4030.	4030.	1.0000000
25-34	Female	NM	12	54016.	19449.	.1296431
25-34	Female	Total	39	177057.	31562.	.0317771
25-34	Total	M	16	77048.	21730.	.0795461
25-34	Total	W	17	75103.	17339.	.0532992
25-34	Total	S	4	35865.	24436.	.4642159
25-34	Total	NM	28	143579.	32466.	.0511296
25-34	Total	Total	65	331593.	42328.	.0162944
35-44	Male	M	14	61855.	15321.	.0613515
35-44	Male	W	1	4392.	4392.	1.0000000
35-44	Male	S	2	8136.	8136.	1.0000000
35-44	Male	NM	9	47179.	16125.	.1168245
35-44	Male	Total	26	121560.	21518.	.0313335
35-44	Female	M	31	136991.	26813.	.0383101
35-44	Female	W	25	105580.	19971.	.0357782
35-44	Female	S	11	49041.	15943.	.1056871
35-44	Female	NM	7	33957.	12997.	.1464932
35-44	Female	Total	74	325569.	43557.	.0178995
35-44	Total	M	45	198846.	30938.	.0242071
35-44	Total	W	26	109972.	20448.	.0345724
35-44	Total	S	13	57176.	17899.	.0979968
35-44	Total	NM	16	81136.	20711.	.0651601
35-44	Total	Total	100	447129.	49484.	.0122478
45-54	Male	M	52	220557.	28133.	.0162699
45-54	Male	W	2	7013.	4964.	.5011174
45-54	Male	S	17	75694.	18987.	.0629197
45-54	Male	NM	12	58138.	17104.	.0865495
45-54	Male	Total	83	361401.	34312.	.0090141
45-54	Female	M	50	210502.	31456.	.0223298
45-54	Female	W	24	102704.	25139.	.0599145
45-54	Female	S	11	46439.	14031.	.0912957
45-54	Female	NM	6	26079.	10685.	.1678766
45-54	Female	Total	91	385723.	37089.	.0092456
45-54	Total	M	102	431059.	48038.	.0124192
45-54	Total	W	26	109717.	26180.	.0569375
45-54	Total	S	13	122132.	23911.	.0383306
45-54	Total	NM	23	84217.	20167.	.0573444

Table I.—Variance estimates for OASDI and SSI participants under SSA-administered programs—Continued

Age	Sex	Marital status[1]	Unweighted count	Estimate	Standard error	Relative variance
			OASDI only—cont.			
45-54	Total	Total	174	747124.	54047.	.0052331
55-64	Male	M	342	1488914.	99257.	.0044441
55-64	Male	W	26	128374.	24778.	.0372551
55-64	Male	S	36	165105.	29969.	.0329479
55-64	Male	NM	17	82124.	21419.	.0680217
55-64	Male	Total	421	1864517.	113389.	.0036984
55-64	Female	M	351	1478573.	93865.	.0040301
55-64	Female	W	202	856463.	63475.	.0054927
55-64	Female	S	41	174779.	28070.	.0257925
55-64	Female	NM	24	103215.	22004.	.0454497
55-64	Female	Total	618	2613029.	120423.	.0021239
55-64	Total	M	693	2967487.	165997.	.0031291
55-64	Total	W	228	984837.	68234.	.0048003
55-64	Total	S	77	339884.	46806.	.0189647
55-64	Total	NM	60	185339.	32915.	.0315395
55-64	Total	Total	1039	4477546.	197917.	.0019538
65-69	Male	M	652	2778693.	145189.	.0027301
65-69	Male	W	38	173900.	31586.	.0329904
65-69	Male	S	42	197829.	30920.	.0244292
65-69	Male	NM	39	178509.	28946.	.0262943
65-69	Male	Total	771	3328931.	158555.	.0022686
65-69	Female	M	603	2445450.	124833.	.0026058
65-69	Female	W	328	1301091.	63726.	.0023989
65-69	Female	S	68	269385.	34190.	.0161081
65-69	Female	NM	53	210263.	35869.	.0291007
65-69	Female	Total	1052	4226188.	146084.	.0011948
65-69	Total	M	1255	5224143.	228339.	.0019104
65-69	Total	W	366	1474991.	73343.	.0024725
65-69	Total	S	110	467214.	48524.	.0107864
65-69	Total	NM	92	388772.	41663.	.0114844
65-69	Total	Total	1823	7555119.	246535.	.0010648
70-74	Male	M	526	2211887.	125904.	.0032400
70-74	Male	W	69	308203.	45817.	.0220994
70-74	Male	S	28	121108.	23433.	.0374377
70-74	Male	NM	27	125257.	24585.	.0385257
70-74	Male	Total	650	2766455.	139422.	.0025399
70-74	Female	M	377	1634980.	104934.	.0041192
70-74	Female	W	379	1626694.	88937.	.0029892
70-74	Female	S	37	162834.	31180.	.0366651
70-74	Female	NM	46	209242.	34337.	.0269301
70-74	Female	Total	839	3633749.	178731.	.0024193
70-74	Total	M	903	3846867.	199390.	.0026865
70-74	Total	W	448	1934897.	107103.	.0030640
70-74	Total	S	65	283942.	37106.	.0170774
70-74	Total	NM	73	334499.	47244.	.0199480
70-74	Total	Total	1489	6400204.	267776.	.0017505
75-84	Male	M	468	1988365.	125679.	.0039952
75-84	Male	W	116	510172.	61289.	.0144324
75-84	Male	S	28	116411.	24034.	.0426257
75-84	Male	NM	22	95184.	15865.	.0277809
75-84	Male	Total	634	2710130.	150989.	.0031039
75-84	Female	M	269	1191177.	84073.	.0049815
75-84	Female	W	585	2679240.	132442.	.0024436
75-84	Female	S	36	160437.	28486.	.0315242
75-84	Female	NM	88	397776.	47085.	.0140117
75-84	Female	Total	978	4428629.	174050.	.0015446
75-84	Total	M	737	3179542.	190234.	.0035797
75-84	Total	W	701	3189411.	153949.	.0023299
75-84	Total	S	64	276848.	36552.	.0174319
75-84	Total	NM	110	492959.	50716.	.0105844
75-84	Total	Total	1612	7138760.	283838.	.0015809
85+	Male	M	57	246861.	32533.	.0173675
85+	Male	W	44	242744.	42750.	.0310149
85+	Male	S	4	18399.	9514.	.2673954
85+	Male	NM	6	35978.	15424.	.1838019
85+	Male	Total	111	543980.	58333.	.0114989
85+	Female	M	25	91970.	17962.	.0381441
85+	Female	W	219	834132.	63100.	.0057225

Table I.—Variance estimates for OASDI and SSI participants under SSA-administered programs—Continued

Age	Sex	Marital status[1]	Unweighted count	Estimate	Standard error	Relative variance
			OASDI only—cont.			
85+	Female	S	7	27090.	10365.	.1463952
85+	Female	NM	10	34102.	10917.	.1024837
85+	Female	Total	261	987293.	71426.	.0052338
85+	Total	M	82	338830.	43377.	.0163895
85+	Total	W	263	1076875.	77735.	.0052107
85+	Total	S	11	45489.	14069.	.0956645
85+	Total	NM	16	70079.	20564.	.0861066
85+	Total	Total	372	1531272.	101393.	.0043844
			SSI only			
Total	Total	Total	335	1550062.	125430.	.0065479
Total	Total	NM	123	546880.	62646.	.0131221
Total	Total	S	80	397264.	43744.	.0121251
Total	Total	W	61	249210.	42864.	.0295840
Total	Total	M	71	356709.	45562.	.0163148
< 18	Male	NM	2	7361.	5246.	.5079297
< 18	Female	NM	1	4370.	4370.	1.0000000
< 18	Total	NM	3	11731.	6828.	.3387586
18-24	Male	NM	13	67973.	20382.	.0899115
18-24	Female	S	1	4271.	4271.	1.0000000
18-24	Female	NM	14	68475.	21556.	.0991042
18-24	Female	Total	15	72745.	21975.	.0912562
18-24	Total	NM	16	136448.	31575.	.0535512
18-24	Total	Total	28	140718.	31863.	.0512712
25-34	Male	M	4	17112.	8626.	.2541071
25-34	Male	NM	9	56268.	19663.	.1221226
25-34	Male	Total	13	73380.	19990.	.0742110
25-34	Female	M	7	30357.	13351.	.1934159
25-34	Female	W	1	2801.	2801.	1.0000000
25-34	Female	S	13	65411.	22161.	.1147831
25-34	Female	NM	21	101224.	24471.	.0584425
25-34	Female	Total	42	199792.	32211.	.0259932
25-34	Total	M	11	47468.	17949.	.1429876
25-34	Total	NM	30	157492.	30640.	.0378502
25-34	Total	Total	55	273171.	36880.	.0182268
35-44	Male	M	2	9521.	6759.	.5040373
35-44	Male	W	1	4726.	4726.	1.0000000
35-44	Male	S	4	20770.	10631.	.2619952
35-44	Male	NM	6	39912.	17092.	.1833900
35-44	Male	Total	13	74928.	23953.	.1021975
35-44	Female	M	7	35734.	13694.	.1468636
35-44	Female	S	18	83043.	21535.	.0672484
35-44	Female	NM	7	32351.	12341.	.1455173
35-44	Female	Total	32	151128.	30387.	.0404283
35-44	Total	M	9	45255.	18444.	.1661076
35-44	Total	S	18	103813.	24016.	.0535186
35-44	Total	NM	13	72262.	21081.	.0851084
35-44	Total	Total	45	226056.	38468.	.0289576
45-54	Male	M	6	27401.	11254.	.1686981
45-54	Male	NM	3	16536.	9654.	.3408730
45-54	Male	Total	9	43936.	14828.	.1138946
45-54	Female	M	11	45134.	12658.	.0786538
45-54	Female	W	5	22396.	10125.	.2044074
45-54	Female	S	17	78309.	16748.	.0457406
45-54	Female	NM	7	32688.	12423.	.1444498
45-54	Female	Total	40	178526.	28290.	.0251104
45-54	Total	M	17	72535.	16176.	.0497318
45-54	Total	NM	10	49223.	15733.	.1021679
45-54	Total	Total	49	222462.	31375.	.0198909
55-64	Male	M	6	27229.	11135.	.1672420
55-64	Male	S	4	22691.	11438.	.2541047
55-64	Male	NM	5	30260.	14131.	.2180876
55-64	Male	Total	15	80179.	20680.	.0665257
55-64	Female	M	9	42124.	16624.	.1557495
55-64	Female	W	10	46112.	14711.	.1017860
55-64	Female	S	17	73164.	15898.	.0472172
55-64	Female	NM	1	5130.	5130.	1.0000000

Table I.—Variance estimates for OASDI and SSI participants under SSA-administered programs—Continued

Age	Sex	Marital status[1]	Unweighted count	Estimate	Standard error	Relative variance
			SSI only—cont.			
55-64	Female	Total	37	166529.	26792.	.0258833
55-64	Total	M	15	69353.	20930.	.0910775
55-64	Total	S	21	95855.	19367.	.0408220
55-64	Total	NM	6	35389.	15033.	.1804569
55-64	Total	Total	52	246708.	35316.	.0204914
65-69	Male	M	6	27450.	13480.	.2411725
65-69	Male	S	1	5738.	5738.	1.0000000
65-69	Male	NM	3	10665.	6212.	.3393365
65-69	Male	Total	10	43852.	15913.	.1316878
65-69	Female	M	6	25670.	10548.	.1688572
65-69	Female	W	10	39949.	13637.	.1165299
65-69	Female	S	4	18963.	9836.	.2690720
65-69	Female	NM	5	19067.	8551.	.2011198
65-69	Female	Total	25	103648.	20832.	.0403968
65-69	Total	M	12	53120.	20067.	.1427083
65-69	Total	S	5	24701.	11388.	.2125446
65-69	Total	NM	8	29731.	10569.	.1263746
65-69	Total	Total	35	147500.	28171.	.0364758
70-74	Male	M	7	26507.	10149.	.1465923
70-74	Male	NM	2	10523.	7442.	.5002612
70-74	Male	Total	9	37030.	12585.	.1155128
70-74	Female	M	3	12172.	7083.	.3386633
70-74	Female	W	6	24366.	9978.	.1677108
70-74	Female	S	3	16302.	9415.	.3335978
70-74	Female	NM	3	12947.	7512.	.3366193
70-74	Female	Total	15	65786.	18699.	.0807925
70-74	Total	M	10	38679.	15046.	.1513221
70-74	Total	NM	5	23470.	10574.	.2030004
70-74	Total	Total	24	102816.	25600.	.0619948
75-84	Male	M	5	19544.	8793.	.2024056
75-84	Male	W	3	8736.	5046.	.3336572
75-84	Male	Total	8	28280.	10138.	.1285093
75-84	Female	M	2	7917.	5598.	.5000312
75-84	Female	W	17	71632.	17733.	.0612834
75-84	Female	S	1	3901.	3901.	1.0000000
75-84	Female	NM	4	23433.	19539.	.6952958
75-84	Female	Total	24	106883.	27254.	.0650218
75-84	Total	M	7	27461.	13089.	.2271973
75-84	Total	W	20	80368.	19766.	.0604910
75-84	Total	Total	32	135163.	33839.	.0626804
85+	Male	S	1	4704.	4704.	1.0000000
85+	Female	M	1	2840.	2840.	1.0000000
85+	Female	W	8	28493.	11111.	.1520652
85+	Female	NM	2	7703.	5467.	.5038455
85+	Female	Total	11	39036.	12705.	.1059296
85+	Total	Total	12	43740.	13548.	.0959363
			OASDI and SSI			
Total	Total	Total	366	1598359.	152132.	.0090592
Total	Total	NM	51	243820.	33439.	.0188086
Total	Total	S	59	259581.	37829.	.0212375
Total	Total	W	168	701867.	69525.	.0098125
Total	Total	M	88	393092.	74110.	.0355438
18-24	Male	NM	2	8441.	5993.	.5040591
18-24	Female	NM	4	18518.	9315.	.2530180
18-24	Total	NM	6	26959.	11076.	.1687959
25-34	Male	S	1	10068.	10068.	1.0000000
25-34	Male	NM	7	33532.	10389.	.0959927
25-34	Male	Total	8	43600.	14467.	.1100987
25-34	Female	W	1	3580.	3580.	1.0000000
25-34	Female	NM	4	17978.	8990.	.2500436
25-34	Female	Total	5	21557.	9676.	.2014712
25-34	Total	NM	11	51510.	13738.	.0711380
25-34	Total	Total	13	65157.	17404.	.0713514
35-44	Male	NM	4	20395.	10223.	.2512503
35-44	Female	W	1	4870.	4870.	1.0000000

Table I.—Variance estimates for OASDI and SSI participants under SSA-administered programs—Continued

Age	Sex	Marital status[1]	Unweighted count	Estimate	Standard error	Relative variance
OASDI and SSI—cont.						
35-44	Female	S	3	11948.	6915.	.3349714
35-44	Female	NM	1	5543.	5543.	1.0000000
35-44	Female	Total	5	22360.	10112.	.2045137
35-44	Total	NM	5	25938.	11629.	.2010072
35-44	Total	Total	9	42755.	14379.	.1131078
45-54	Male	M	1	6263.	6263.	1.0000000
45-54	Male	W	1	4059.	4059.	1.0000000
45-54	Male	S	1	5157.	5157.	1.0000000
45-54	Male	NM	5	25960.	13638.	.2759768
45-54	Male	Total	8	41439.	16379.	.1562345
45-54	Female	M	1	3789.	3789.	1.0000000
45-54	Female	W	1	4022.	4022.	1.0000000
45-54	Female	S	6	31886.	13127.	.1694950
45-54	Female	NM	2	8454.	5995.	.5028209
45-54	Female	Total	10	48150.	15453.	.1029962
45-54	Total	M	2	10052.	7320.	.5302876
45-54	Total	W	2	8080.	5713.	.5000105
45-54	Total	S	11	37043.	11909.	.1033558
45-54	Total	NM	10	34414.	14897.	.1873841
45-54	Total	Total	18	89589.	22334.	.0621503
55-64	Male	M	6	25913.	12198.	.2215964
55-64	Male	W	1	4987.	4987.	1.0000000
55-64	Male	S	2	10625.	7717.	.5276068
55-64	Male	NM	3	15120.	8737.	.3339024
55-64	Male	Total	12	56643.	17594.	.0964783
55-64	Female	M	8	38486.	14040.	.1330844
55-64	Female	W	11	46099.	14788.	.1029058
55-64	Female	S	9	34385.	12596.	.1341939
55-64	Female	NM	2	9177.	6489.	.5000067
55-64	Female	Total	30	128146.	23980.	.0350169
55-64	Total	M	14	64399.	20216.	.0985467
55-64	Total	W	12	51085.	16923.	.1097383
55-64	Total	S	12	45010.	14772.	.1077163
55-64	Total	NM	5	24296.	10883.	.2006359
55-64	Total	Total	42	184789.	32842.	.0315871
65-69	Male	M	12	53931.	17970.	.1110225
65-69	Male	W	2	7523.	5437.	.5222957
65-69	Male	S	1	6603.	6603.	1.0000000
65-69	Male	Total	15	68057.	18906.	.0771726
65-69	Female	M	6	24831.	8618.	.1204502
65-69	Female	W	32	129568.	26794.	.0427633
65-69	Female	S	5	22668.	10161.	.2009360
65-69	Female	NM	3	12794.	7440.	.3382045
65-69	Female	Total	46	189861.	29768.	.0245832
65-69	Total	M	18	78762.	22078.	.0785764
65-69	Total	W	34	137091.	29934.	.0476783
65-69	Total	S	6	29271.	12118.	.1713932
65-69	Total	Total	61	257917.	37955.	.0216558
70-74	Male	M	8	31406.	10147.	.1043939
70-74	Male	W	3	11621.	6777.	.3401275
70-74	Male	S	2	8966.	6391.	.5080770
70-74	Male	NM	3	15018.	8770.	.3410458
70-74	Male	Total	16	67010.	20146.	.0903885
70-74	Female	M	11	50253.	17738.	.1245843
70-74	Female	W	39	163619.	29621.	.0327747
70-74	Female	S	13	54596.	15206.	.0775686
70-74	Female	NM	4	16552.	8410.	.2581784
70-74	Female	Total	67	285020.	43907.	.0237310
70-74	Total	M	19	81659.	21322.	.0681784
70-74	Total	W	42	175240.	31815.	.0329614
70-74	Total	S	15	63562.	16201.	.0649655
70-74	Total	NM	7	31570.	12151.	.1481469
70-74	Total	Total	83	352029.	51120.	.0210879
75-84	Male	M	19	83750.	27374.	.1068347
75-84	Male	W	8	39519.	14358.	.1320007
75-84	Male	S	3	11340.	6703.	.3494227
75-84	Male	NM	1	4216.	4216.	1.0000000
75-84	Male	Total	31	138824.	30551.	.0484303

Table I.—Variance estimates for OASDI and SSI participants under SSA-administered programs—Continued

Age	Sex	Marital status[1]	Unweighted count	Estimate	Standard error	Relative variance
			OASDI and SSI—cont.			
75-84	Female	M	11	49022.	15289.	.0972771
75-84	Female	W	37	163484.	28646.	.0307027
75-84	Female	S	8	34864.	12522.	.1289945
75-84	Female	NM	6	24888.	8451.	.1153048
75-84	Female	Total	62	272257.	39936.	.0215161
75-84	Total	M	30	132771.	39096.	.0867091
75-84	Total	W	45	203003.	32494.	.0256211
75-84	Total	S	11	46204.	14203.	.0944941
75-84	Total	NM	7	29103.	9444.	.1053015
75-84	Total	Total	93	411081.	58833.	.0204830
85+	Male	M	3	15476.	5219.	.1137502
85+	Male	W	5	22409.	10090.	.2027545
85+	Male	S	1	6166.	6166.	1.0000000
85+	Male	Total	9	44050.	12925.	.0861010
85+	Female	M	2	9975.	7060.	.5010565
85+	Female	W	26	96512.	17763.	.0338759
85+	Female	S	2	10312.	7744.	.5639426
85+	Female	NM	2	7238.	5122.	.5007861
85+	Female	Total	32	124036.	22002.	.0314648
85+	Total	M	5	25450.	8883.	.1218274
85+	Total	W	31	118920.	20795.	.0305792
85+	Total	S	3	16477.	9898.	.3608795
85+	Total	Total	41	168085.	26407.	.0246826

[1] NM = Never married; S = Separated; W = Widowed; M = Married.

Chart I.—Half-sample assignment for pseudo-unit 1 cases

Stratum	Half-sample
1	11111110111010011011100011010110100011101001010011100010011010001000000
2	11111101110100110111000110101101000111010010100111000100110100010000001
3	11111011101001101110001101011010001110100101001110001001101000100000011
4	11110111010011011100011010110100011101001010011100010011010001000000111
5	11101110100110111000110101101000111010010100111000100110100010000001111
6	11011101001101110001101011010001110100101001110001001101000100000011111
7	10111010011011100011010110100011101001010011100010011010001000000111111
8	01110100110111000110101101000111010010100111000100110100010000001111111
9	11101001101110001101011010001110100101001110001001101000100000011111110
10	11010011011100011010110100011101001010011100010011010001000000111111101
11	10100110111000110101101000111010010100111000100110100010000001111111011
12	01001101110001101011010001110100101001110001001101000100000011111110111
.	
.	
.	
66	00000011111110111010011011100011010110100011101001010011100010011010001
67	00000111111101110100110111000110101101000111010010100111000100110100010
68	00001111111011101001101110001101011010001110100101001110001001101000100
69	00011111110111010011011100011010110100011101001010011100010011010001000
70	00111111101110100110111000110101101000111010010100111000100110100010000
71	01111111011101001101110001101011010001110100101001110001001101000100000

NOTE 2: Evaluation of Direct Variance Estimates from the 1984 SIPP Public Use File

Case weights and variable values were based on the rotation group as shown below:

Rotation group	Month(*)
1	4
2	3
3	2
4	1

The variables are referred to by their public use file names. (Starting character position of the month-1 field is shown in parentheses.)

1. Age 16 and over
   AGE-*(2206)>16.

2. Low Income Cash Only (LICO)
   H*TOTINC(178)<H*POV$(173).

3. LICO plus government noncash transfers (LICNC)
   H*TOTINC(178)+H*NONCSH(215)<H*POV(173).

4. Receiving Unemployment Compensation (UNCO)
   IO5AMT*(3820)+IO6AMT*(3848)+IO7AMT*(3876)>0.

5. Receiving Cash from a means tested program (CBPR)
   H*-TRAN(201)>0.

6. Receiving food stamps (FS)
   H*-FDSTP(251)>0.

7. Receiving noncash benefits other than food stamps (NCBPR)
   CAIDCOV*(2672)=1, or
   H*PUBAMT(258)>0, or
   H*-LUNCH(266)≠0, or
   H*-BREAK(267)≠, or
   H*-4804(269)>0, or
   H*NONSCH(215)>H*-FDSTP(251).

8. Some labor force activity (SLFA)
   ESR-*(2593)≥1, and
   ESR-*(2593)≤7.

9. Hispanic (HIS)
   ETHNICTY(2278)≥14, and
   ETHNICTY(2278)≤20.

Evaluation of Direct Variance Estimates from the 1984 SIPP Public Use File

INTRODUCTION

The 1984 public use data files of the Survey of Income and Program Participation (SIPP) provide pseudo stratum and pseudo primary sampling unit codes that permit direct estimates of sampling variances by a number of methods. The actual sample design parameters are withheld from public use to prevent access to small geographic areas where disclosure of individual identities might be possible. The Social Security Administration (SSA) has used the pseudo codes to compute sampling variances for SSA program participants. (Bye and Gallicchio 1988.) Although the variance estimates appeared to be reasonably well behaved, no external assessment of them was made.

In this note we report the results of a comparison of direct variance estimates from the public use file with variance estimates based on the original sample design computed by the Bureau of the Census. The comparison involves estimates of 36 population totals that comprised the item set for the first generalized curve ("program participation and benefits, poverty") in the SIPP User's Guide (1987, page 7-5). The SSA direct variance estimates were computed using 72 balanced half samples derived from the pseudo design. Details are provided in Bye and Gallicchio (1988). The Census estimates were obtained from a set of 50 half samples that were not fully balanced derived from the original design. Case weights in each of the Census half samples were adjusted to a common set of population totals, replicating the weighting methodology of the full sample. The SSA half sample case weights were constructed by multiplying the full sample weight by 2.

The results of the comparison are very encouraging. Most of the items compared showed small differences in coefficient of variation (CV). The differences were both positive and negative with no apparent pattern. This finding together with the ease of computation of the estimator makes the direct estimation of variances from the public use sample very attractive to the data analyst.

VARIANCE ITEMS

This section presents the SSA item specifications. (An exact match of public use file estimates with those provided by Census was not expected because the Census estimates were produced some years ago from an internal file for which specifications are not longer available.) The 36 items were combinations of 9 characteristics (Bureau of the Census, 1985). SSA's construction of these characteristics relate to individual and household status as of September 1983.

RESULTS

Table 1 presents the comparison of Census and SSA variance estimates for the 36 items. As expected the estimated totals do not agree exactly, and these differences contribute to the differences in estimated standard errors. A more meaningful comparison, therefore, is the ratio of CVs. With the exception of items 26 and 32, the Census and SSA CVs are quite similar. The ratios of the SSA CV to Census CV range from a low of .849 to a high of 1.093. There is no apparent pattern to the differences as a function of size of the estimate.

The SSA CV for item 26 (item 32 consists of essentially the same sample cases as 26) is about 50 percent larger than the corresponding Census CV. An examination of the 72 SSA half sample estimates of this characteristics (data not shown here) indicates a wide range of estimated totals but no extreme outliers. The size of the CV for this estimate appears to be a chance occurrence indicating, perhaps, that the SSA variance estimator might have a larger variance than the Census estimator, especially when cells are small. A comparison of substantially more items would be needed to investigate this further.

REFERENCES

Bye, Barry V. and Gallicchio, Salvator J., "A Note on Sampling Variance Estimates for Social Security Program Participants From the Survey of Income and Program Participation," Social Security Bulletin, Vol. 51 No. 10, October 1988.

Survey of Income and Program Participation, User's Guide, Bureau of the Census, Department of Commerce, July 1987.

Memorandum for Documentation from Karen E. King, Subject: SIPP Variances: Items for Generalized Variance Parameters, Bureau of the Census, Department of Commerce, June 15, 1985.

Table 1. Census/SSA Variance Comparison

	Census			SSA			Ratio SSA/Census		
	Estimate	St. Err.	CV	Estimate	St. Err.	CV	Estimate	St. Err.	CV
1. low income cash (lico) only	20875350	693897	0.033	21459181	707972	0.033	1.028	1.020	0.993
2. lico plus govt noncash trans (licnc)	19791310	684505	0.035	20373140	701869	0.034	1.029	1.025	0.996
3. receiving unemp. compensation (unco)	3038883	154221	0.051	2916786	156861	0.054	0.960	1.017	1.060
4. receiving cash from mean-test (cbpr)	14556060	703225	0.048	14546133	651057	0.045	0.999	0.926	0.926
5. receiving food stamps (fs)	11920830	551678	0.046	12006268	535354	0.045	1.007	0.970	0.964
6. receiving noncash o/t fs (ncbpr)	25730990	875913	0.034	24921228	898371	0.036	0.969	1.026	1.059
7. some labor force activity (slfa)	112328300	1884640	0.017	115087192	1901174	0.017	1.025	1.009	0.985
8. slfa in hh with licnc	9497841	324529	0.034	9625030	354010	0.037	1.013	1.091	1.076
9.slfa receiving unco	2913951	149957	0.051	2799895	155409	0.056	0.961	1.036	1.079
10.slfa in hh receiving fs	5516297	258996	0.047	5469103	246196	0.045	0.991	0.951	0.959
11.no labor force act (nlfa) in hh lico	10841920	445140	0.041	11296831	455121	0.040	1.042	1.022	0.981
12.nlfa receiving unco	124932	24073	0.193	116890	23229	0.199	0.936	0.965	1.031
13.nlfa in hh receiving cbpr	8674453	432411	0.050	8810027	434336	0.049	1.016	1.004	0.989
14.same as 3. in hh with lico	519714	66145	0.127	556148	64921	0.117	1.070	0.981	0.917
15.same as 4. in hh with lico	6949680	450128	0.065	6899747	412336	0.060	0.993	0.916	0.923
16.same as 5. in hh with lico	8561362	454857	0.053	8514647	431563	0.051	0.995	0.949	0.954
17.same as 6. in hh with lico	11244680	507620	0.045	10558880	500328	0.047	0.939	0.986	1.050
18.same as 7. in hh with lico	10033130	323310	0.032	10162351	358123	0.035	1.013	1.108	1.094
19.same as 9. in hh with lico	502881	63081	0.125	539965	64644	0.120	1.074	1.025	0.954
20.same as 10. in hh with lico	3628493	208210	0.057	3485122	187039	0.054	0.960	0.898	0.935
21.same as 12. in hh with lico	16833	8519	0.506	16183	8105	0.501	0.961	0.951	0.990
22.same as 13. in hh with lico	4588350	306016	0.067	4686842	310529	0.066	1.021	1.015	0.993
23. number of hispanics 16+	9606953	625095	0.065	9461055	651284	0.069	0.985	1.042	1.058
24. his low income cash (lico) only	2371371	274453	0.116	2247679	275540	0.123	0.948	1.004	1.059
25.his lico plus govt noncash trans (licnc)	2212253	258185	0.117	2112670	262870	0.124	0.955	1.018	1.066
26. his receiving unemp. compensation (unco)	235565	35162	0.149	240340	52543	0.219	1.020	1.494	1.465
27. his receiving cash from mean-test (cbpr)	1704630	212180	0.124	1669796	203091	0.122	0.980	0.957	0.977
28. his receiving food stamps (fs)	1602333	210802	0.132	1555791	187125	0.120	0.971	0.888	0.914
29. his receiving noncash o/t fs (ncbpr)	3090985	316105	0.102	3116432	330621	0.106	1.008	1.046	1.037
30. his some labor force activity (slfa)	6231358	385368	0.062	6201126	408599	0.066	0.995	1.060	1.065
31. his slfa in hh with licnc	1008518	123187	0.122	988474	130064	0.132	0.980	1.056	1.077
32.his slfa receiving unco	218963	32774	0.150	227275	51143	0.225	1.038	1.560	1.503
33.his slfa in hh receiving fs	672542	109441	0.163	682383	94327	0.138	1.015	0.862	0.849
34. his no labor force act (nlfa) in hh lico	1296891	161290	0.124	1202075	154376	0.128	0.927	0.957	1.033
35.his nlfa receiving unco	16596	8850	0.533	13065	7613	0.583	0.787	0.860	1.093
36.his nlfa in hh receiving cbpr	1054751	126333	0.120	1000704	128222	0.128	0.949	1.015	1.070