Appendix C. Source and Accuracy of Estimates

SOURCE OF DATA

The data were collected during the sixth wave of the 1985 panel and the third wave of the 1986 panel of the Survey of Income and Program Participation (SIPP). The SIPP universe is the noninstitutionalized resident population of persons living in the United States.[1]

The 1985 and 1986 panel SIPP samples are located in 230 Primary Sampling Units (PSUs) each consisting of a county or a group of contiguous counties. Within these PSUs, expected clusters of 2 to 4 living quarters (LQs), were systematically selected from lists of addresses prepared for the 1980 decennial census to form the bulk of the sample. To account for LQs built within each of the sample areas after the 1980 census, a sample was drawn of permits issued for construction of residential LQs up until shortly before the beginning of the panel. In jurisdictions that do not issue building permits, small land areas were sampled and the LQs within were listed by field personnel and then subsampled. In addition, sample LQs were selected from a supplemental frame that included LQs identified as missed in the 1980 census.

Approximately 17,800 and 16,300 LQs were originally designated for the 1985 and 1986 samples, respectively. For the first interview period, interviews were obtained from the occupants of about 13,400 of the 17,800 designated LQs for the 1985 panel and 11,500 of the 16,300 designated LQs for the 1986 panel. Most of the remaining 4,400 and 4,800 LQs in the 1985 and 1986 panels respectively, were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, approximately 1,000 of the 4,400 LQs for the 1985 panel, and 900 of the 4,800 LQs for the 1986 panel were not interviewed because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or were otherwise unavailable. Thus, occupants of about 92 percent of all eligible living quarters participated in the first interview of each panel.

For the subsequent interviews, only original sample persons (those interviewed at the first contact) and persons living with them were eligible to be interviewed. With certain restrictions, original sample persons were to be followed if they moved to a new address. When original sample persons moved without leaving forwarding addresses, moved to remote parts of the country and no telephone number was available, or refused to be interviewed, additional noninterviews resulted.

Sample households within a given panel are divided into four subsamples of nearly equal size. These subsamples are called rotation groups 1, 2, 3, or 4, and one rotation group is interviewed each month. Each household in the sample was scheduled to be interviewed at 4 month intervals over a period of roughly 2 1/2 years beginning in February of 1985 for the 1985 panel and February of 1986 for the 1986 panel. The reference period for the questions is the 4-month period preceding the interview month. In general, one cycle of four interviews covering the entire sample, using the same questionnaire, is called a wave. The exceptions are Wave 2 of the 1985 panel and Wave 3 of the 1986 panel which cover three interviews.

As part of most waves, subjects are covered that do not require repeated measurement during the panel and are of particular interest cross-sectionally for research purposes. A specific set of topical questions are referred to as a topical module. For this report the topical modules analyzed include questions on health status and utilization of health care services and long-term care. They were implemented in Wave 6 of the 1985 panel and Wave 3 of the 1986 panel.

Since Wave 6 of the 1985 panel and Wave 3 of the 1986 panel are concurrent and contain the same relevant topical modules on health and disability, the data were combined and analyzed as a single data set. The primary motivation for combining this data is to obtain an increase in sample size in conjunction with a reduction in time in sample bias, if any, due to nonresponse over the life of a panel.

Noninterviews. Tabulations in this report were drawn from interviews conducted from September through December 1986. Table C-1 summarizes information on nonresponse for the interview months in which the data used to produce this report were collected.

[1]The noninstitutionalized resident population includes persons living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates and nursing home residents, were not eligible to be in the survey. Also, United States citizens residing abroad were not eligible to be in the survey. With these qualifications, persons who were at least 15 years of age at the time of interview were eligible to be interviewed.

Table C-1. **Combined 1985, 1986 Household Sample Size by Month and Interview Status**

Month	Eligible	Inter-viewed	Nonin-ter-viewed	Nonre-sponse rate (%)
September 1986	3300	2700	600	*18
October 1986	6700	5500	1200	18
November 1986	6600	5500	1200	18
December 1986	6600	5400	1200	18

* Due to rounding of all numbers at 100, there are some inconsistencies. The percentage was calculated using unrounded numbers.

Some respondents do not respond to some of the questions. Therefore, the overall nonresponse rate for some items such as income and money related items is higher than the nonresponse rates in table C-1.

ESTIMATION

The estimation procedure used to derive SIPP person weights involved several stages of weight adjustments. In the first wave, each person received a base weight equal to the inverse of their probability of selection. For each subsequent interview, each person received a base weight that accounted for following movers.

A noninterview adjustment factor was applied to the weight of every occupant of interviewed households to account for households which were eligible for the sample but were not interviewed. (Individual nonresponse within partially interviewed households was treated with imputation. No special adjustment was made for noninterviews in group quarters.) A factor was applied to each interviewed person's weight to account for the SIPP sample areas not having the same population distribution as the strata from which they were selected. The Bureau uses complex techniques to adjust the weights for nonresponse, but the success of these techniques in avoiding bias is unknown.

An additional stage of adjustment to persons' weights was performed to reduce the mean square errors of the sample estimates by ratio adjusting SIPP sample estimates to monthly Current Population Survey (CPS) estimates[2] of the civilian (and some military) noninstitutional population of the United States by age, race, sex, type of householder (married, single with relatives, single without relatives), and relationship to householder (spouse or other). The CPS estimates were themselves brought into agreement with estimates from the 1980 decennial census which were adjusted to reflect births, deaths, immigration, emigration, and changes in the Armed Forces since 1980. Also, an adjustment was made so that a husband and wife within the same household were assigned equal weights.

[2]These special CPS estimates are slightly different from the published monthly CPS estimates. The differences arise from forcing counts of husbands to agree with counts of wives.

ACCURACY OF ESTIMATES

SIPP estimates in this report are based on a sample; they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: nonsampling and sampling. The magnitude of SIPP sampling error can be estimated, but this is not true of nonsampling error. Found below are descriptions of sources of SIPP nonsampling error, followed by a discussion of sampling error, its estimation, and its use in data analysis.

Nonsampling variability. Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness on the part of the respondents to provide correct information, inability to recall information, errors made in collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, biases resulting from the differing recall periods caused by the rotation pattern used, and failure to represent all units within the universe (undercoverage). Quality control and edit procedures were used to reduce errors made by respondents, coders, and interviewers.

Undercoverage in SIPP results from missed living quarters and missed persons within sample households. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-blacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have different characteristics than the interviewed persons in the same age-race-sex group. Further, the independent population controls used have not been adjusted for undercoverage in the decennial census.

Unique to the 1986 panel, maximum telephone interviewing was tested in Waves 2, 3, and 4. Specifically, half of the sample in rotations 4 and 1 of Wave 2 and rotations 2 and 3 of Wave 3 (Phase I) and rotations 2, 3, and 4 of Wave 4 (Phase II) were designated for telephone interviews. Analysis (done by designated mode) of household nonresponse, item nonresponse rates for labor force and income core items, and selected cross-sectional estimates of recipiency, income, low income status, and selected topical module items gave no indication of an overall significant mode effect. However, analysis was restricted to a limited number and type of estimates. If differences between two time periods or differences in characteristics for demographic groups result in borderline significant differences, the significance may be due to bias from the use of the telephone mode. Similarly, borderline insignificant differences may also be due to this bias. Thus, although no overall significant mode effect was detected, the user should consider the possibility of mode effects while analyzing combined data involving the 1986 panel after Wave 1, especially results based on Waves 2 through 4 data. Details on analyses are in "Preliminary Evaluation of Maximum Telephone Interviewing on the SIPP" (paper by Gbur and Petroni in the forthcoming 1989 *Proceedings of the Survey Research Methods Section, American Statistical Association)* and "SIPP 86: Telephone Experiment - Preliminary Analysis" (internal Census Bureau draft memorandum from Waite to Davey, August 21, 1989).

Comparability with other estimates. Caution should be exercised when comparing data from this report with data from earlier SIPP publications or with data from other surveys. The comparability problems are caused by sources such as the seasonal patterns for many characteristics, different nonsampling errors, and by different concepts and procedures in other surveys.

Sampling variability. Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

USES AND COMPUTATION OF STANDARD ERRORS

Confidence intervals. The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

Hypothesis testing. Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common types of hypotheses tested are 1) the population parameters are identical versus 2) they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the parameters are different when, in fact, they are identical.

All statements of comparison in the report have passed a hypothesis test at the 0.10 level of significance or better. Therefore, for most differences cited in the report, the estimated absolute difference between parameters is greater than 1.6 times the standard error of the difference.

To perform the most common test, compute the difference X_A - X_B, where X_A and X_B are sample estimates of the parameters of interest. A later section explains how to derive an estimate of the standard error of the difference X_A - X_B. Let that standard error be s_{DIFF}. If X_A - X_B is between -1.6 times s_{DIFF} and +1.6 times s_{DIFF}, no conclusion about the parameters is justified at the 10 percent significance level. If, on the other hand, X_A - X_B is smaller than -1.6 times s_{DIFF} or larger than +1.6 times s_{DIFF}, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the parameters are different. Of course, sometimes this conclusion will be wrong. When the parameters are, in fact, the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed more erroneous significant differences will occur. For example, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, the significance of any single test should be interpreted cautiously.

Note when using small estimates. Summary measures (such as percent distributions) are shown in the report only when the base is 200,000 or greater. Because of the large standard errors involved, there is little chance that summary measures would reveal useful information when computed on bases smaller than

200,000. Also, nonsampling error in one or more of the small number of cases providing the estimate can cause large relative error in that particular estimate. Estimated numbers are shown, however, even though the relative standard errors of these numbers are larger than those for the corresponding percentages. These smaller estimates are provided primarily to permit such combinations of the categories as serve each user's needs. Therefore, care must be taken in the interpretation of small differences since even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Standard error parameters and tables and their use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because clusters of living quarters are sampled for SIPP. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required. Estimates with similar standard error behavior were grouped together and two parameters (denoted "a" and "b") were developed to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These "a" and "b" parameters vary by characteristic and by demographic subgroup to which the estimate applies. Table C-4 provides base "a" and "b" parameters to be used for estimates in this report.

Table C-2. **Standard Errors of Estimated Numbers of Persons**

(Numbers in thousands)

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200	59	50,000	821
300	72	80,000	949
600	102	100,000	988
1,000	131	130,000	988
2,000	185	135,000	982
5,000	290	150,000	950
8,000	365	200,000	676
11,000	425	220,000	417
13,000	460	230,000	99
15,000	491		
17,000	521		
22,000	585		
26,000	630		
30,000	670		

For those users who wish further simplification, we have also provided general standard errors in tables C-2 and C-3. Note that these standard errors must be adjusted by an "f" factor from table C-4. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors are given in the following sections.

Standard errors of estimated numbers. The approximate standard error, s_x, of an estimated number of persons shown in this report can be obtained in two ways.

It may be obtained by use of the formula

$$s_x = fs \qquad (1)$$

where f is the appropriate "f" factor from table C-4, and s is the standard error on the estimate obtained by interpolation from table C-2. Alternatively, s_x may be approximated by the formula

$$s_x = \sqrt{ax^2 + bx} \qquad (2)$$

from which the standard errors in table C-2 were calculated. Here x is the size of the estimate and "a" and "b" are the parameters associated with the particular type of characteristic being estimated. Use of formula (2) will provide more accurate results than the use of formula (1) above.

Illustration. SIPP estimates from table D of this report show that 8,548,000 persons age 15 to 64 received food stamps during the 1-month period prior to their interview month. The appropriate "a" and "b" parameters and "f" factor from table C-4 and the appropriate general standard error found by interpolation from table C-2 are

$$a = -0.0000816,\ b = 13,892,\ f = 0.90,\ s = 376,000$$

Using formula (1), the approximate standard error is

$$0.90 \times 376,000 = 338,000$$

The 90-percent confidence interval as shown by the data is from 8,007,200 to 9,088,800. Using formula (2), the approximate standard error is

$$\sqrt{(-0.0000816)(8,548,000)^2 + (13,892)(8,548,000)} = 336,000.$$

The 90-percent confidence interval as shown by the data is from 8,010,400 to 9,085,600. Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all samples.

Standard errors of estimated percentages. This section refers to the type of percentages presented in this report. These are the percentages of a group of persons possessing a particular attribute. An example of this type of percentage would be the percentage of persons needing assistance with personal care. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are over 50 percent. For example, the percentage of persons needing assistance is more reliable than the estimated number of persons needing assistance. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

For the percentage of persons, the approximate standard error, $s_{(x,p)}$, of the estimated percentage p can be obtained by the formula

$$s_{(x,p)} = fs. \quad (3)$$

Table C-3. **Standard Errors of Estimated Percentages of Persons**

Base of Estimated Percentage (Thousands)	Estimated Percentage					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	2.9	4.1	6.4	8.8	12.7	14.7
300	2.4	3.4	5.2	7.2	10.4	12.0
600	1.7	2.4	3.7	5.1	7.3	8.5
1,000	1.3	1.8	2.9	3.9	5.7	6.6
2,000	1.0	1.3	2.0	2.8	4.0	4.6
5,000	0.6	0.8	1.3	1.8	2.5	2.9
8,000	0.5	0.7	1.0	1.4	2.0	2.3
11,000	0.4	0.6	0.9	1.2	1.7	2.0
13,000	0.4	0.5	0.8	1.1	1.6	1.8
17,000	0.3	0.5	0.7	1.0	1.4	1.6
22,000	0.3	0.4	0.6	0.8	1.2	1.4
26,000	0.3	0.4	0.6	0.8	1.1	1.3
30,000	0.2	0.3	0.5	0.7	1.0	1.2
50,000	0.2	0.3	0.4	0.6	0.8	0.9
80,000	0.2	0.2	0.3	0.4	0.6	0.7
100,000	0.1	0.2	0.3	0.4	0.6	0.7
130,000	0.1	0.2	0.3	0.4	0.5	0.6
180,000	0.1	0.1	0.2	0.3	0.4	0.5
200,000	0.1	0.1	0.2	0.3	0.4	0.5
230,000	0.1	0.1	0.2	0.3	0.4	0.4
250,000	0.1	0.1	0.2	0.3	0.4	0.4

In this formula, f is the appropriate "f" factor from table C-4 and s is the standard error of the estimate obtained by interpolation from table C-3. Alternatively, $s_{(x,p)}$ may be approximated by the formula

$$s_{(x,p)} = \sqrt{\left(\frac{b}{x}\right) p\,(100-p)} \quad (4)$$

from which the standard errors in table C-3 were calculated. Here x is the size of the subclass of persons which is the base of the percentage, p is the percentage ($0<p<100$), and b is the "b" parameter associated with the characteristic in the numerator. Use of this formula formula will give more accurate results than use of formula (3) above.

Table C-4. **SIPP Generalized Variance Parameters for 1985, 1986 Combined Panel**

Characteristics[1]	a	b	f
PERSONS			
Total or White			
Health and Disability (4)	-0.0000027	7,802	0.67
16+ Program Participation and Benefits, Poverty (5)			
Both Sexes	-0.0000816	13,892	0.90
Male	-0.0001716	13,892	
Female	-0.0001554	13,892	
16+ Income and Labor Force (6)			
Both Sexes	-0.0000278	4,736	0.52
Male	-0.0000586	4,736	
Female	-0.0000530	4,736	
All Others[2] (7)			
Both Sexes	-0.0000747	17,224	1.00
Male	-0.0001545	17,224	
Female	-0.0001446	17,224	
Black			
Health and Disability (1)	-0.0000027	7,802	0.67
Poverty (2)			
Both Sexes	-0.0004265	11,850	0.83
Male	-0.0009103	11,850	
Female	-0.0008023	11,850	
All Others (3)			
Both Sexes	-0.0002310	6,372	0.61
Male	-0.0004963	6,372	
Female	-0.0004319	6,372	

[1]For cross-tabulations, use the parameters of the characteristic with the smaller number within the parentheses.

[2]Use the "All Others" parameters for retirement tabulations, 0+ program participation, 0+ benefits, 0+ income, and 0+ labor force tabulations, in addition to any other types of tabulations not specifically covered by another characteristic in this table.

Illustration. Table 2 in the report shows that during the 4-month period prior to their interview month an estimated 11.8 percent of persons 15 years and over with monthly household income under $600 need assistance with one or more activities. [The base of 16,227,000 person for this percentage is given in table 1 of the report.] Using formula (3) with the "f" factor from table

C-4 and the appropriate standard error from table C-3, the standard error is

$$s_{(x,p)} = 0.67 \times 1.1\% = 0.7\%.$$

Using formula (4) with the "b" parameter from table C-4, the approximate standard error is

$$s_{(x,p)} = \sqrt{\frac{7,802}{16,227,000} 11.8\%(100\% - 11.8\%)} = 0.7\% .$$

Consequently, the 90-percent confidence interval as shown by these data is from 10.7 to 12.9 percent.

Standard error of a difference. The standard error of a difference between two sample estimates is approximately equal to

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2 - 2rs_xs_y} \qquad (5)$$

where s_x and s_y are the standard errors of the estimates x and y, and r is the correlation coefficient between the characteristics estimated by x and y. The estimates can be numbers, percents, ratios, etc. Underestimates or overestimates of standard errors of differences result if the estimated correlation coefficient is overestimated or underestimated, respectively.

Illustration. Suppose we are interested in the difference in the percentage of persons 15 years and over with monthly household income under $600 and $3,000 and over who need assistance with one or more activities. Of the 16,227,000 persons 15 years and over with monthly household income under $600 (see table 1 of report) and the 67,509,000 persons 15 years and over with monthly household income $3,000 and over (see table 1 of report), 11.8 percent and 1.9 percent, respectively, needed assistance with one or more activities (see table 2 of report). Using formula 5 and the "b" parameter from table C-4, the standard errors of these percentages are approximately 0.7 percent and 0.1 percent, respectively.

Now, the standard error of the difference is computed using the above two standard errors. The correlation between these estimates is assumed to be zero. Therefore, the standard error of the difference is computed using formula (5):

$$s_{(x-y)} = \sqrt{(0.7)^2 + (0.1)^2} = 0.7 \textit{ percent}$$

Suppose that it is desired to test at the 10 percent significance level whether the two percentages differ significantly. To perform the test compare the difference of 9.9 percent to the product 1.6 x 0.7 percent = 1.12 percent. Since the difference is larger than 1.6 times the standard error of the difference, the data show that the estimates of 11.8 and 1.9 percent differ significantly at the 10 percent level.