

ADJUSTING FOR COVERAGE ERRORS IN THE MARCH 1973 CURRENT POPULATION SURVEY

Denton R. Vaughan
Social Security Administration
and
C. Terrence Ireland
The George Washington University

This paper presents a progress report on methods for adjusting coverage errors in the Current Population Survey (CPS). The work which will be discussed here is part of a larger effort being mounted jointly by the Census Bureau and the Social Security Administration (SSA) with the assistance of the Internal Revenue Service (IRS). The project involves a match of the March 1973 CPS to Social Security administrative records and to limited income information from Federal income tax returns.^{1/} As part of the reconciliation of the coverage differences between these sources, it is necessary to "correct" for the CPS's understatement of the number of persons in the United States civilian noninstitutional population.

Some of the understatement in the CPS (between 3 and 4 percent) is eliminated routinely by the Census Bureau. This is done by inflating the sample to independent estimates of the current population calculated by carrying forward 1970 Census figures to account for subsequent aging of the population, births, deaths, and migration between the United States and other countries.^{2/} However, this inflation process does not correct for the underenumeration known to exist in the 1970 Census itself. Jacob Seigel of the Census Bureau, for example, has estimated that about 5.3 million persons were missed in 1970.^{3/}

PRELIMINARY CPS COVERAGE ADJUSTMENT METHOD FOR MARCH 1973

In their policy simulation work with the Current Population Survey, SSA researchers have found it convenient to group persons in CPS-sampled households into what we will call "dependency units."^{4/} These units are made up of persons usually considered interdependent under social insurance and other income transfer programs. They can consist of a single adult 22 years of age or older, an adult with minor children (under age 14), and married couples with or without minor children. Young adults (age 14 to 21), depending on their living arrangements, are treated as separate units or as part of a unit containing their parent(s). Unmarried adults age 22 or older without dependents are always treated as separate units regardless of their living arrangements.

Adjustment Procedure.--An essential requirement for employing the dependency unit scheme in this facet of SSA's policy research is for all members of the same unit to have the same sample weight. Consequently, it is important that the adjustment procedure assign a single weight to all members of a unit, as well as make the sample consistent with overall corrected controls.

One way to incorporate the Census undercount corrected totals into the sample would be to follow the same procedure used in the second-stage weighting of the CPS itself. In the Census Bureau's weighting of the CPS, independent age-race-sex-specific population controls are incorporated by separately adjusting every person in the sample by factors which depend on the individual's age, race, and sex.^{5/} This practice, while leading to substantial overall improvements in the survey's principal estimates,^{6/} does have certain unpleasant side effects. Inconsistencies arise, for example, between the count of the number of men estimated to be living with their wives and the number of women estimated to be living with their husbands. (In fact, since the basic sampling unit in the CPS is the household, any difference which exists between these counts is due wholly to the way the control totals are introduced.)

SSA's requirement of a single weight for each unit makes the simple ratio estimation step described above unsuitable. The method we chose for making the adjustment was described in a paper given at last year's annual meetings.^{7/} The technique, known as "Multivariate Raking," has a number of attractive features. While iterative, it generally converges rapidly; also, the estimators obtained have optimal statistical properties under simple random sampling. The CPS, of course, is a complex multi-stage, stratified, cluster sample; but, as was shown last year, the performance of the multivariate raking algorithm was quite satisfactory with the CPS (achieving variance reductions in over half the cases examined).

The "corrected" controls for the U.S. civilian noninstitutional population were obtained with the help of Bob Warren and Gary Shapiro of the Census Bureau.

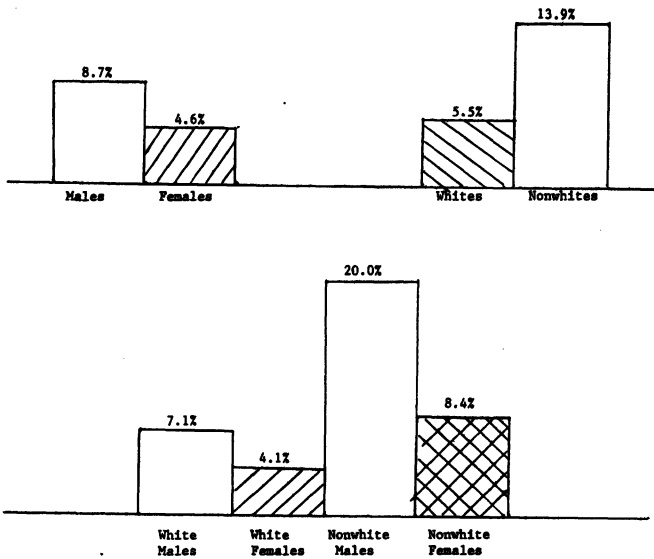
UNDERCOVERAGE CHARACTERISTICS

Before looking at the impact of the correction on selected dependency unit statistics, it might be of interest to examine the overall CPS coverage problem by age, race, and sex.

Race and Sex.--Figure 1 shows the total March 1973 CPS undercoverage for persons age 14 or older by race and sex as a percent of the "true," or corrected, population. It illustrates the differences between the undercoverage of males and females, whites and nonwhites, and among the four race-sex subgroups. Looking at the top half of the chart, you will notice that a much larger percentage of males was missed than females -- 8.7 percent versus 4.6 percent; and more nonwhites were missed than whites -- 13.9 percent versus 5.5

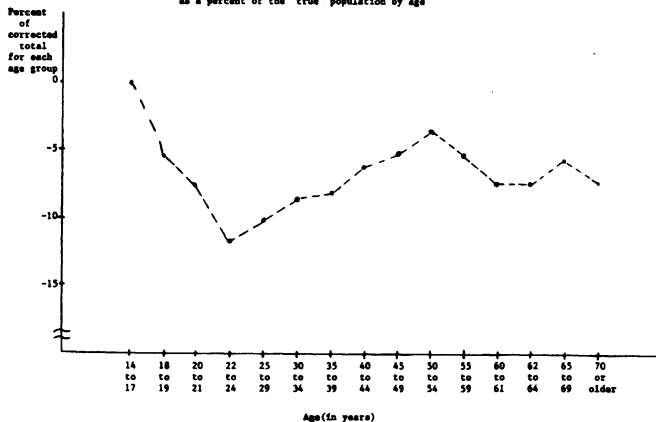
percent. Of the four race-sex subgroups, the largest percentage undercoverage was 20.0 percent for nonwhite males.

Figure 1--Total March 1973 CPS undercoverage for persons age 14 or older as a percent of the "true" population by race and sex



Age.--Figure 2 shows the total March 1973 CPS undercoverage for persons age 14 or older by age as a percent of the "true" population. As can be seen, the largest percentage undercoverage is about 12 percent for the 22 to 24 age group. Of course, there is considerable variation in the extent of undercoverage by age among the four race-sex subgroups which is not reflected in the figure. In general, this variation is analogous to the variation in the aggregate undercoverage rates by subgroup. Thus, among persons under 50, age-specific undercoverage rates for white females were usually less than half the age-specific rates of the total population. The undercount rates for nonwhite males, on the other hand, were about two to four times greater, age for age, than those for the population as a whole at ages below 60.

Figure 2--Total March 1973 CPS undercoverage for persons age 14 or older as a percent of the "true" population by age



Other Factors.--Post-enumeration surveys following the 1950 and 1960 Decennial Censuses provide clues

to other aspects of the coverage problem, in addition to age, race, and sex. For example, they indicate that persons with loose attachments to households, such as lodgers and boarders, were more likely to be missed than household heads, wives, or children.^{8/} The chance, for example, of missing a primary family unit with young children is believed to have been less than that of missing a young adult male without dependents, even though all the individuals happen to be living together at the time of the interview.

These coverage differences within and between households can also be thought of as translating roughly into a similar set of differential coverage probabilities by type of dependency unit. On that basis, we introduced an overall control on the number of households, in an attempt, naive as it turns out, to correct for these differences.

Tabulations from the 1950 Post-Enumeration Survey (PES) comparing the occupation and income of males enumerated in the census with those missed in the census, but picked up in the PES, also indicate that the occupational status and income of missed persons was lower than the enumerated population's even after standardizing by color.^{9/}

The adjustment process has two kinds of impacts. The first will result directly from the correction of the sample controls; i.e., differential but predetermined shifts in the age, sex, and race components of the population. The second type of impact affects the sample's other variables and is introduced indirectly through the manipulation of the controls. Changes of the first kind are, of course, largely straightforward. Changes of the second type are more problematical and must be assessed closely to determine if the adjustment process has yielded sensible results. Of key interest from the standpoint of the present project is the adjustment's impact on the distribution of the kinds of dependency units and on the array of variables related to socioeconomic status (SES). To the extent that the coverage biases associated with the decennial censuses and the CPS are similar, we believe that it is reasonable to expect a drop in the relative importance of larger units (especially primary families with children) and to see small declines in SES within each race-sex subgroup.

PRELIMINARY ADJUSTMENT'S IMPACT BY RACE-SEX SUBGROUP

Our review of the adjustment's effects will concentrate on the changes produced within the four major race-sex subgroups. The impact of the procedure by sex and, especially, race were frequently greater than the within race-sex group changes. However, they were largely only a direct by-product of a combination of the altered controls and, in the case of nonwhites, for example, substantial within race-sex differentials in unit characteristics and SES.

Focusing on these within group changes, table 1 shows the total number of dependency units before

and after adjustment for CPS coverage errors. Overall, the number of dependency units increased 5.7 percent. There was a considerably larger increase in the number of nonwhite-headed units than was so for white-headed units. In fact, the largest change shown in table 1 was the 29 percent increase in the number of dependency units that were headed by nonwhite males. There was a larger increase in the number of units without children than for those with children. For female heads, increases in the number of units without children were also accompanied by actual decreases in the number of units with children. This tendency was slightly more marked for units headed by nonwhite females than white females. For nonwhite males, the increase was slightly larger for those with children; just opposite the pattern observed among the other three race-sex subgroups.

Table 1--Number of dependency units before and after adjustment for CPS coverage errors, units with and without children separately

Race and sex of unit head	(number in thousands)								
	ALL UNITS			UNITS WITH CHILDREN			UNITS WITHOUT CHILDREN		
	Number after adjustment	Number before adjustment	Percent increase	Number after adjustment	Number before adjustment	Percent increase	Number after adjustment	Number before adjustment	Percent increase
Total	87,131	82,453	5.67	34,021	32,541	4.55	53,110	49,912	6.41
Male	63,942	59,731	7.05	29,266	27,674	5.75	34,676	32,058	8.17
Female	23,189	22,722	2.06	4,755	4,868	-2.32	18,434	17,854	3.25
WHITE									
Total	76,059	73,032	4.14	29,287	28,483	2.82	46,772	44,549	4.99
Male	56,811	54,205	4.81	26,158	25,292	3.42	30,652	28,913	6.01
Female	19,249	18,826	2.25	3,128	3,191	-1.97	16,120	15,635	3.10
NONWHITE									
Total	11,072	9,422	17.51	4,734	4,058	16.66	6,337	5,363	18.16
Male	7,131	5,526	29.04	3,107	2,382	30.44	4,024	3,144	27.99
Female	3,940	3,896	1.13	1,627	1,676	-2.92	2,314	2,219	4.28

Note: Children are defined as all dependent persons (i.e., not heads or wives of unit heads) who are under 22 years of age. The percentages shown here are in terms of the "uncorrected" estimates not the "corrected" ones as is true in figures 1 and 2.

Considering the nature of the missed population, the changes in the number of dependency units by race and sex resulting from the correction procedure are not surprising. With the exception of nonwhite male-headed units, the relative size and direction of the changes in the number of units with and without children also seem reasonable. As units with children are perhaps the most likely to have been interviewed by the enumerators, the pattern exhibited by nonwhite male-headed units in this regard is not as expected.

Unit Size.--The median size of dependency units decreased slightly or remained the same for each of the race-sex subgroups except nonwhite units headed by males. Although the increase was small for this subgroup (0.8 percent), it parallels the increase in the proportion of nonwhite male-headed units with children. It should be noted that this shift did not conform to our expectations, since larger units probably have higher coverage rates, on the average, than smaller ones. (See table 2.)

Unit Type.--Changes in the distribution of units by type, depicted in table 3, also seem sensible, with declines in the proportion of primary family units occurring in three of the four race-sex subgroups. Increases occurred in the proportion of units accounted for by unrelated individuals and other one-person units. Again, it was the units headed by nonwhite males that exhibited an atypical pattern. They showed a slight increase in the proportion of units classified as primary

families. While not shown in the tables, another fact worth noting is that these increases were greater for nonwhite male-headed primary families with children than for those without.

Age and Socioeconomic Status.--The median age of male unit heads tended to decrease while that of female heads increased somewhat. The data in tables 2 and 3 indicate that a decrease in the median age of male heads was accompanied by an increase in educational attainment and the proportion reporting professional and technical occupations; it also resulted in a decrease in the proportion who reported that they had never worked. Shifts experienced by female-headed units in terms of these variables were either less pronounced or in the opposite direction from male heads of the same race: nonwhite male heads demonstrated the greatest increases, while the declines of nonwhite female heads were proportionately greater than those experienced by white female heads. Changes in the earnings distribution of dependency units were slight, and they largely conformed to the pattern of change within race-sex subgroups among the other variables. The percentage of units reporting one or more earners increased more among male-headed than female-headed units. The largest increase was for nonwhite male-headed units; the only decrease occurred for nonwhite female-headed units. Median earnings increased for nonwhite male-headed units. However, among whites, the median for male-headed units was virtually unaffected, while median earnings for female-headed units increased slightly.

In line with this sex-age pattern, the proportion of units reporting at least one member with social security benefits decreased for male-headed units and increased for female-headed units (table 3). Again, the changes were magnified somewhat by race. In terms of the size distribution of unit social security income, there was a slight downward shift in the median for units headed by whites, regardless of sex (table 2). Among units headed by nonwhites, social security income tended to increase slightly for units with male heads and decrease slightly for units with female heads.

Summary.--Overall, the size and direction of the adjustment's effect by race and sex is as expected. Furthermore, the direction of change in the number of units by presence of children, unit size, and type of unit, also, seems reasonable for each of the four race-sex subgroups except nonwhite males. For this group, the changes were opposite the direction expected, with increases in units with dependent children, in the proportion of primary family units, and in unit size.

In our opinion, the pattern of change in dependency unit socioeconomic status by subgroup did not yield sensible results in all cases. Unit SES tended to decline consistently only among nonwhite units headed by females. For nonwhite male-headed units, SES clearly seems to have increased. Whether the SES of units headed by whites of each sex actually increased is unclear, but the data certainly do not indicate that it declined.

The coverage adjustment just described provides a consistent weighting of the March 1973 Current Population Survey to age-race-sex control totals corrected for the 1970 Census undercount. We expect these weights to be of value to researchers working with the 1973 Match Project data tapes that are now being made available as a public-use file.^{10/} However, we do not consider the approach taken here to be a definitive one, as important coverage biases undoubtedly remain. In fact, we are more than a little dissatisfied with our results. While, by the very nature of the problem, some of the CPS coverage bias can never be eliminated, there does seem to be a good bit more that could be done using information we already possess from Census Bureau studies of the undercount. At this point, it appears to us that one of the reasons for the partial failure of the adjustment procedure stems from the fact that the control we included on the number of households was an overall one. Separate controls by race and sex of the household head, if they had been available, might have improved our results considerably.

Using what we have learned from the shortcomings of the present study, together with previous experience at SSA in introducing household controls to correct for the CPS coverage problem,^{11/} we plan to redo the adjustment. Essentially the approach will expand on a procedure employed by the Census Bureau in its March Supplement weighting of the CPS^{12/} combined with the multivariate raking techniques which underlie the methods described in this paper.

- 1/ For some general background on this project's goals and procedures, see the papers in these Proceedings that were given at the session on "Reconciliation of Survey and Administrative Income Distribution Statistics through Data Linkage."
- 2/ For a description of this methodology, see U.S. Bureau of the Census, Current Population Reports, Series P-25, No. 519.
- 3/ This is Seigel's "preferred" estimate of the total missed. (See Estimate of Coverage of Population by Sex, Race and Age: Demographic Analysis, PHC(E)-4, U.S. Bureau of the Census, 1974.)
- 4/ For a detailed discussion of the derivation of dependency units see Mary P. Johnston's Technical Appendix in Projector, Dorothy S., Mary T. Millea, and Kenneth M. Dymond, "Projection of March Current Population Survey: Population, Earnings, and Property Income, March 1972 to March 1976," Studies in Income Distribution, No. 1. DHEW Publication (SSA) 75-11776, 1974. (In that appendix dependency units are referred to as "STATS" units.)
- 5/ U.S. Bureau of the Census, Technical Paper No. 7, Government Printing Office, Washington, D.C., 1963.
- 6/ Banks, M. and Shapiro, G., "Variance of the Current Population Survey, including within and between-PSU components and the effect of the different stages of estimation," 1971 Proc. Amer. Stat. Assn. Soc. Stat. Sec., pp. 40-49, 1972.
- 7/ Ireland, C.T., and Fritz Scheuren, "The Rake's progress" An invited paper presented before the 1974 Annual meetings of the Biometric Society.
- 8/ Johnston, Denis F., and James R. Wetzel, "Effect of the census undercount on labor force estimates," Monthly Labor Review, March 1969, pp. 11-12.
- 9/ See for example Pritzker, L. and N. Rothwell, "Procedural difficulties in taking past censuses in predominantly Negro, Puerto Rican, and Mexican areas," Social Statistics and the City, Cambridge, Harvard University, 1968, pp. 55-79.
- 10/ For documentation on this file see Studies in Interagency Data Linkages, Reports nos. 4, 5 and 6, Social Security Administration, 1975.
- 11/ Scheuren, Fritz, Beth Kilss and H. Lock Oh, "Coverage differences, noninterview nonresponse, and the 1960 Census undercount: 1963 Pilot Link Study," Studies from Interagency Data Linkages, DHEW Pub. No. (SSA) 75-11750, December 1973.
- 12/ Hanson, R. H. and Turner, A. G. March CPS weighting for the 1107. Unpublished Census Bureau Memorandum from Joseph Waksberg to Herman Miller and Daniel Levine, 1967.

TABLE 2.--MEDIAN VALUES OF SELECTED CHARACTERISTICS OF DEPENDENCY UNITS AS ESTIMATED FROM THE MARCH 1973 CPS BEFORE AND AFTER ADJUSTMENT FOR UNDERCOVERAGE BY RACE AND SEX OF UNIT HEAD

ITEM	Total			White			Nonwhite		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
UNIT SIZE									
Before adjustment.....	2.48	2.87	1.64	2.50	2.89	1.60	2.32	2.67	1.88
After adjustment.....	2.47	2.85	1.63	2.49	2.87	1.60	2.35	2.69	1.85
Percent change <u>a</u> /.....	-0.40	-0.70	-0.61	-0.40	-0.69	0.00	1.29	0.75	-1.60
AGE OF UNIT HEAD(in years)									
Before adjustment.....	45.05	43.33	52.65	45.57	43.52	55.40	41.17	41.49	40.59
After adjustment.....	44.55	42.64	52.93	45.20	42.99	55.44	40.38	39.90	41.56
Percent change <u>a</u> /.....	-1.11	-1.59	0.53	-0.81	-1.22	0.80	-1.92	-3.83	2.39
HIGHEST GRADE COMPLETED OF UNIT HEAD									
Before adjustment.....	12.30	12.37	12.11	12.35	12.41	12.18	11.11	11.30	10.89
After adjustment.....	12.30	12.37	12.11	12.36	12.43	12.19	11.24	11.51	10.81
Percent change <u>a</u> /.....	0.00	0.00	0.00	0.08	0.16	0.08	1.17	1.86	-0.73
UNIT'S TOTAL EARNINGS (in dollars)									
Before adjustment <u>b</u> /....	8,681	10,220	3,959	9,104	10,491	4,142	5,660	7,352	3,105
After adjustment <u>a</u> /.....	8,713	10,170	3,980	9,147	10,493	4,171	5,978	7,443	3,065
Percent change <u>a</u> /.....	0.37	-0.49	0.53	0.47	0.02	0.70	5.62	1.24	-1.29
UNIT'S SOCIAL SECURITY INCOME (in dollars)									
Before adjustment <u>b</u> /....	1,823	2,307	1,487	1,863	2,364	1,511	1,467	1,708	1,244
After adjustment <u>b</u> /.....	1,817	2,300	1,484	1,859	2,361	1,510	1,467	1,711	1,232
Percent change <u>a</u> /.....	-0.33	-0.30	-0.20	-0.21	-0.13	-0.07	0.00	0.18	-0.96

a/ Adjusted minus unadjusted divided by unadjusted times one hundred.

b/ Medians calculated excluding units with zero earnings or zero social security income.

TABLE 3.--SELECTED CHARACTERISTICS OF DEPENDENCY UNITS AS ESTIMATED FROM THE MARCH 1973 CPS BEFORE AND AFTER ADJUSTMENT FOR UNDERCOVERAGE BY RACE AND SEX OF UNIT HEAD

ITEM	Total			White			Nonwhite		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Total <u>a/</u>	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
UNIT TYPE									
Before adjustment:									
Primary family.....	60.59	76.79	18.00	61.69	78.09	14.47	52.02	63.96	35.10
Unrelated individuals...	19.56	10.47	43.41	19.33	9.74	46.94	21.27	17.70	26.34
Other 1 person units <u>b/</u> .	18.09	11.59	35.20	17.55	11.10	36.13	22.28	16.35	30.70
Other.....	1.76	1.15	3.39	1.43	1.07	2.47	4.41	1.99	7.87
After adjustment:									
Primary family.....	60.55	76.26	17.24	61.62	77.79	13.90	53.23	64.13	33.52
Unrelated individuals...	19.48	10.64	43.86	19.28	9.81	47.26	20.86	17.32	27.26
Other 1 person units <u>b/</u> .	18.24	11.92	35.66	17.70	11.33	36.50	21.91	16.60	31.53
Other.....	1.72	1.17	3.25	1.39	1.07	2.33	4.00	1.96	7.69
Percent change: <u>c/</u>									
Primary family.....	-0.07	-0.69	-4.22	-0.11	-0.38	-3.94	2.33	0.27	-4.50
Unrelated individuals...	-0.41	1.62	1.04	-0.26	0.72	0.68	-1.93	-2.15	3.49
Other 1 person units <u>b/</u> .	0.83	2.85	1.31	0.85	2.07	1.02	-1.66	1.53	-5.94
Other.....	-2.27	1.74	-4.13	-2.80	0.00	-9.72	-9.30	-1.51	-2.29
OCCUPATION OF LONGEST JOB OF UNIT HEAD IN 1972									
Before adjustment:									
Never worked.....	23.09	14.42	45.89	22.18	13.98	45.80	30.14	18.73	46.32
Professional and technical.....	10.52	11.57	7.76	11.04	12.03	8.18	6.49	7.03	5.71
In Armed Forces.....	1.12	1.54	0.00	1.17	1.57	0.00	0.75	1.28	0.00
Other.....	65.27	72.47	46.35	65.61	72.42	46.02	62.62	72.96	47.97
After adjustment:									
Never worked.....	22.57	14.10	45.94	21.80	13.70	45.72	27.87	17.31	46.97
Professional and technical.....	10.60	11.63	7.76	11.15	12.15	8.21	6.80	7.49	5.54
In Armed Forces.....	1.15	1.56	0.00	1.19	1.59	0.00	0.86	1.34	0.00
Other.....	65.68	72.71	46.32	65.86	72.56	46.07	64.47	73.86	47.49
Percent change: <u>c/</u>									
Never worked.....	-2.25	-2.22	0.11	-1.71	-2.00	-0.17	-7.53	-7.58	1.62
Professional and technical.....	0.76	0.52	0.00	1.00	1.00	0.37	11.66	6.54	-2.98
In Armed Forces.....	2.68	1.30	0.00	1.71	1.27	0.00	14.67	4.69	0.00
Other.....	0.63	0.89	-0.11	0.38	0.19	0.11	2.95	1.23	1.11
UNITS WITH ONE OR MORE EARNERS									
Before adjustment.....	79.14	88.13	55.51	79.95	88.48	55.41	72.84	84.71	56.00
After adjustment.....	79.63	88.42	55.41	80.31	88.74	55.44	74.98	85.90	55.23
Percent change <u>c/</u>	0.62	0.33	-0.18	0.45	0.29	0.05	2.94	1.40	-1.38
UNITS WITH ONE OR MORE SOCIAL SECURITY RECIPIENTS									
Before adjustment.....	21.72	16.36	35.82	22.28	16.48	38.98	17.37	15.14	20.53
After adjustment.....	21.16	15.75	36.10	21.89	16.06	39.08	16.19	13.23	21.53
Percent change <u>c/</u>	-2.58	-3.73	0.78	-1.75	-2.55	0.26	-6.79	-12.62	4.87

a/ Percents may not add to 100 due to rounding.

b/ The category "Other one person units" is synonymous with the category "Individuals-from-a-family" (See the source cited in the fourth footnote at the end of this paper.)

c/ Adjusted minus unadjusted divided by unadjusted times one hundred.