Income in the American Community Survey: Comparisons to Census 2000[1]

Kirby G. Posey, Edward Welniak, and Charles Nelson
HHES Division, U.S. Census Bureau, Washington, DC 20233

The long-term goal of the American Community Survey (ACS) is to produce more timely local area estimates. Currently, the only sources of Census Bureau local area income estimates are the decennial census (data only once every ten years) and the Small Area Income and Poverty Estimates (SAIPE) program (annual county and state estimates of median household income). The Census Bureau designed the battery of income questions as well as the edit and allocation specifications for the ACS to be as similar as possible to that of the Census in order to evaluate the ACS's performance. There are, of course, some differences inherent with the ACS that forced some changes. The purpose of this paper is to explain the major differences and provide a preliminary overview of the ACS results. This paper builds upon and updates a previous Census Bureau paper on this subject (Posey and Welniak 1999)

In Census 2000, a battery of eight detailed income questions and a total income question were posed to all persons 15 years old and over. The eight detailed income questions included (1) wages and salary, (2) self-employment income, (3) interest, dividend, and net royalty income, (4) Social Security, (5) Supplemental Security Income (SSI), (6) public assistance or welfare, (7) retirement, survivor, or disability income, and (8) any other sources of income received on a regular basis. These eight income questions were placed at the end of each set of person questions on the long form (sample) questionnaire so that response levels to the other less sensitive sample questions would not be adversely impacted. A "total income" question was first asked in the 1980 Census to aid in the resolution of income entry problems in the detailed questions. Both the 1990 and 2000 Census long form questionnaires also included the "total income" question. Test results indicated that a sizable number of all income recipients furnished responses to the total income question, but failed to provide answers to some or all parts of the detailed type of income questions, proving the total income item very important.

The biggest difference between collection methods in the ACS and Census 2000 is the income reference period. The ACS collects data throughout the year on an on-going, monthly basis. The ACS asks for a respondent's income over the "past 12 months." The census, however, collected the income data for a fixed period of time – "during 1999" (the last calendar year.)

The Split Panel Test

To evaluate the impact of this reference period change, the Census Bureau conducted a "split panel" test of about 19,000 households (with a 49 percent mail response rate). This split panel test was conducted from October through December of 1997. This special survey collected income statistics using both terminologies, "past 12 months" and "income in calendar year 1996" for the two different randomly assigned treatment groups. There were a few limitations to the test that should be noted. First, the test consisted of the mail responses only; no follow-up interviews were conducted. Also, the data were not processed or edited.

The only statistical differences in median income estimates between the two reference periods occurred in the earnings categories, wages / salary and self-employment (see Table 1).

It had been expected that the "past 12-months" reference period would yield slightly higher estimates because of the more recent reference period. As shown in Table 1, that is not always the case. For example, the median wage/salary income reported on forms asking about the PAST 12 MONTHS was $25,000 for individuals. The median wage/salary income reported on forms asking about the PAST CALENDAR YEAR is $26,000. For self-employment income, however, the "past 12 months" reference period yielded the higher figure. All other income items did not differ between the two reference periods.

When looking at response rates (the percentage of respondents who answered "yes," "no," or "loss") by the different sources of income, the questionnaire with the "past 12 months" reference period produced slightly higher response rates for every income source although only one income

[1]This paper reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

item, public assistance, showed a statistically significant difference over the two reference periods. (See Table 2.)

Because response rates were slightly better and to assure that the income questions would be as consistent as possible with the other, more current information collected on the ACS (disability and labor force status, for example), the Census Bureau decided to keep the 12-month reference period for the ACS.

Making Adjustments for the 12 Reference Periods

Since the American Community Survey interviews households monthly and asks for income received during the "past twelve months," it ends up with twelve different reference periods for income data for a single calendar year's worth of data collection (see Figure 1). That is, if a respondent is asked in February 2000 how much income they received during the past 12 months, their reference period was February 1999 through January 2000. If, on the other hand, a respondent is interviewed in October 2000 and asked about income during the past twelve months, the reference period would then be October 1999 through September 2000. When data in the ACS profiles and the Summary Files are released on the Internet, they are referred to as "2000 inflation-adjusted dollars." That is, the income data are presented for a fixed reference period. To accomplish this, adjustment factors are applied to the data depending on the month of interview to adjust everyone's income to constant 2000 dollars. The Consumer Price Index (CPI-U) is used to do this. (See Figure 2.)

Respondents who were interviewed in February 2000 and asked about their income from February 1999 through January 2000 had their income adjusted by a factor of 1.031447. Similarly, respondents who were interviewed in October and whose income reference period was from October 1999 through September 2000 had their incomes adjusted by a factor of 1.008443. These factors were computed as the sum of the 2000 CPI monthly adjustment factors divided by the sum of the CPI monthly adjustment factors for the reference period. The result is that all ACS income data collected in 2000 is presented in terms of consistent 2000 calendar-year dollars. Note that on the public use files, for disclosure avoidance purposes, there is only one adjustment factor for all respondents. This is to keep from disclosing the month of interview to help protect the individual confidentiality of our respondents. That factor is the average of the 12 monthly factors.

Also note that the ACS is not designed to produce income statistics for use on a monthly basis. Data are collected over a 12-month period and then aggregated to an annual amount.

ACS (C2SS) and Census 2000 Edits and Allocations: Similarities and Differences

The Census 2000 Supplementary Survey (C2SS) was an ACS program designed to demonstrate the feasibility of collecting long Census form-type information at the same time but separately from the decennial census. C2SS used the ACS questionnaire and methodology to collect demographic, social, economic, and housing data from a sample of around 700,000 households nationwide. Much of the analysis in the rest of this paper is based on comparisons between the C2SS and Census 2000 at the national and state level.

For the most part, ACS (and therefore the C2SS) used the same editing scheme that was used for Census 2000. For example, in the consistency edits for both surveys, obvious mistakes such as monthly amounts (rather than annual) reported for income sources like social security, public assistance or retirement income were checked and corrected, as were the common mistakes respondents sometimes make with decimal points. An example of this kind of correction would be someone who reported $100,000 in public assistance income where they should have reported $1,000. Another edit checked for respondents who might have intended to report a "loss" in self-employment income or in interest, dividends, or net rental income but forgot to mark the "loss" box . Also, respondents often confuse the first income item for the total income field and report their total income twice. The "total income" field was used to resolve differences in the reporting of income components when possible. Where a correction was not obvious, no changes were made.

A major difference between the two surveys was in the method used for data capture. In Census 2000, most of the enumerator-filled and mail-return questionnaires were processed using Optical Character Recognition, or OCR, an image scanning system in which interpretive software is applied to digital images of the questionnaire fields. The ACS/C2SS did not employ this system of data capture. The ACS/C2SS employed actual "keyers" during the data capture operation. In the census, OCR interpreted numeric handwritten income entries then performed data quality checks to help ensure that the number read was accurate. If the entry failed to meet an accuracy threshold, an image of the item in question was displayed to a keyer who then entered the response. This basic difference in data capture methodology resulted in some modifications to the Census 2000 edits.

For missing income data, the design of the allocation and imputation specifications for the C2SS/ACS were basically the same as those used in Census 2000. In both surveys, files were first sorted by sex. That is, reported data (hot deck values) were stored separately for males and females so that missing economic characteristics are allocated to a male only from another male; never from a female, and vice versa. Then, for each sex, reported data were further sorted and stored according to race and ethnicity, educational attainment, living arrangements (husband or wife, other family reference person, other family members and unrelated individuals), and finally by residence (whether a person did or did not live in a metropolitan area). For example, if a respondent failed to report earnings data but did provide answers to other items such as occupation, class of worker, weeks worked last week, and age, their reported data were matched against that of another respondent who fully reported all items. When a match was located, the fully reported earnings data were substituted for the missing information.

In fact, both the ACS and Census 2000 employed a two-level imputation system. That is, an attempt was made to find a similar respondent based on a detailed matrix of characteristics, hoping to match as closely as possible a donor (with reported information) to a recipient (with missing information). If no match is found, a more general and less detailed matrix (or second match "key") was used to impute the missing information. This type of procedure is particularly appropriate for surveys like the C2SS, where sample size is smaller than that of the decennial long form, and this smaller sample size may discourage the use of extremely detailed donor/recipient match characteristics.

Sometimes respondents will indicate they received a particular type of income but will not report how much they received. The ACS and Census 2000 treated respondents who reported income "recipiency" with no amount differently than those respondents who left an income question completely blank. That is, missing amounts were allocated from different matrices depending on whether the respondent marked the "yes" box for that item. In other words, an attempt is made to establish a pattern of recipiency and then go from there. The allocation process in both surveys was basically the same.

The biggest difference between the allocation processes between the two surveys was the sample size. Once the ACS is fully implemented, the results of the ACS allocation procedure will more closely mirror that of the Census because the larger sample size will enhance the imputation system's ability to locate a better match for non-respondents.

Examining the Results: C2SS/ACS and Census 2000 Median Incomes

As the intention of the ACS program is to replace the long form in 2010 as the major source of small area income survey data, it is important to compare the results of the C2SS to those of the 2000 Census. The comparisons below are between those based on ACS/C2SS interviews conducted in calendar year 2000 and results from Census 2000. State and national estimates are compared and data from the Current Population Survey (CPS) are included as well for comparison purposes.

Median household income estimates were generally lower in the C2SS/ACS than Census 2000 after adjusting the 1999 dollar values from Census 2000 for inflation. (See Table 3.)

Median household income for the nation was more than 4 percent lower in C2SS than in Census 2000. Median household incomes in Arizona, Louisiana, Nebraska, Nevada, and Oregon were more than 8 percent lower in the C2SS than in the census. Only three states in the C2SS had median household income estimates slightly higher than that of the Census: Hawaii, Mississippi, and Rhode Island, none of which statistically significant.

C2SS median household income at the national level matched up more closely with that of the Current Population Survey (CPS) estimates. Table 3 also compares the C2SS results with calendar-year 2000 estimates from the CPS (based on Census 2000-consistent population controls). The CPS national median household income estimate was $41,990. The C2SS estimate, $41,486, was only 1.2 percent lower. Still that difference was significant. The C2SS estimates of median household incomes were significantly lower than those of the CPS in 13 states. The C2SS estimates came in significantly higher than those of the CPS in 8 states.

It is interesting (and somewhat puzzling) to note that the Census 2000 estimated median household income at the national level in 2000 dollars ($43,396) was 3.2 percent higher than the calendar-year 2000 estimate from the CPS. This is puzzling because conventional wisdom tells us that the CPS, which has a much more detailed set of income questions than Census 2000/C2SS/ACS, would result in higher income levels. Thus, the fact that the C2SS median estimate at the national level is slightly lower than the CPS estimate is not surprising. The fact that the Census 2000 median is quite a bit higher than the CPS is, however, quite surprising. Of the three major Census Bureau

household survey-based estimates of median incomes at the national level, the outlier is the Census 2000 estimate, not the C2SS or CPS estimate. This relationship is generally true at the state level as well. For example, the Census 2000 estimate of median household income in California was $49,078, about 5 percent higher than both the CPS and C2SS estimates. The difference between decennial and Current Population Survey income estimates is not solely a Census 2000 phenomenon. The estimate of U.S. median household income was $30,056 from the 1990 Census, 4.3 percent higher than the comparable figure from the March 1990 CPS ($28,820).

Exploring Possible Reasons for C2SS/ACS and Census 2000 Median Income Differences

We looked into several possible explanations for the higher income amounts in the Census. The difference could be the result of certain census/ACS differences such as survey design, mode of interviewing, different data capture/processing methods, as well as the different reference periods.

As far as the reference periods are concerned there is one possible explanation for the lower figures for the "PAST 12 MONTHS." A respondent might read the instructions and "key in" on the word "Month" and report "monthly" amounts in the income fields. The income edits are designed to catch monthly amounts and would have caught many of these types of mistakes but some of the higher monthly amounts indicated by respondents could have remained unchanged, thus lowering the median income. As a precautionary measure and to help prevent this situation, the questionnaire instructions were changed slightly in the late 1990s to emphasize annual amounts. The instructions were changed to read "Mark (X) the "Yes" box for each type of income this person received, and give your best estimate of the TOTAL AMOUNT during the PAST 12 MONTHS. (NOTE: The "past 12 months" is the period from today's date one year ago up through today.)" The 2000 C2SS/ACS questionnaire reflected this change although monthly income amounts still may be at least partially to blame for the lower median household incomes.

Another possible reason for the higher income amounts in Census 2000 is the data capture methodology. As previously stated, the Census Bureau used the Optical Character Recognition (OCR) technology to read handwritten data entry by machine in the census, whereas ACS employs keyers to record this information. If OCR errors tend to add to incomes, on balance, more than keying errors, (for example, OCR might be more inclined to read a dollar sign as an 8 than a keyer), that tendency may result in somewhat higher Census 2000 income estimates. However, as the CPS/1990 and 2000 Census comparisons indicate, national median household incomes were higher in both the 1990 Census (which did not employ OCR technology) and Census 2000 (which did).

When making comparisons between the Census income figures and those of the C2SS/ACS, it is also important to keep in mind the different modes of data collection. That is, when gathering information from households who failed to respond via mail, the C2SS/ACS used computer-assisted telephone interviewing (CATI) and computer-assisted personal interviewing (CAPI) whereas Census 2000 used personal interviews. This is yet another possible explanation for the differences in median household income.

Also, as previously discussed, the 1997 split-panel test showed that respondents report wage or salary income significantly lower when asked about the "past 12 months" than they do when asked about the last calendar year. Wage or salary income makes up most (about 75 percent) of aggregate household income. Further analysis is needed in order to determine why respondents report lower wage/salary amounts for the "past 12 months" versus the last calendar year.

In addition, it is clear that the inflation adjustment procedure for income (the procedures described previously that use Consumer Price Index (CPI) data to convert C2SS/ACS dollar amounts into 2000 calendar-year dollars) also play a role when comparing C2SS/ACS data with other data sources. As explained earlier, C2SS/ACS reference periods vary by month of interview; the average factor used to convert C2SS/ACS dollar amounts into 2000 calendar-year dollars was approximately 1.019. In order to make the comparisons between C2SS/ACS and Census 2000 income estimates that were shown in table 3, the Census 2000 estimates had to be converted from 1999 calendar-year dollars into 2000 calendar-year dollars. That factor was approximately 1.033. If no CPI adjustment had been made to the dollars reported on either Census 2000 or C2SS/ACS, the difference between medians at the U.S. level would have been smaller than the 4.6 percent shown in Table 3 (2.5 percent). Since adjustment clearly played a role in determining the size of the difference between Census 2000 and C2SS/ACS estimates, it would be worthwhile to examine the costs and benefits of adjusting C2SS/ACS incomes as well as the choice of factors used to adjust them.

There is also the possibility that allocation differences play a role, though much more work

would have to be done to understand how allocation and differences in allocation methods (as well as differences in sample size, which has a direct impact on the size of the allocation "donor pool") affected the final results. One of the reasons to look to allocation difference as a potential source of ACS/Census 2000 income differences is the fact that both of these surveys have relatively high income allocation rates. For Census 2000, around 30 percent of respondents had at least some income items allocated. For the C2SS, the comparable allocation rate was around 24 percent.

Conclusions and Avenues for Further Research

This paper presented income comparisons between the Census 2000 Supplementary Survey (C2SS) and Census 2000 at the national and state level. In doing so, it also provided a summary of the major differences between the two income data sources, in terms of data collection, capture, and processing, and provided very preliminary assessments of the possible role these differences may have played. CPS estimates of income at the national and state level were also provided in order to put the C2SS/Census 2000 differences in perspective.

The most obvious limitation to this study is that we don't know the "truth," meaning that we don't know the true income of the C2SS and Census 2000 survey respondents. Thus, it is impossible to make firm judgments about the relative data quality of these quite different sources of sub-national survey-based income estimates.

The work presented here is very preliminary; there is much that needs to be done as we make the transition from the Census long form to the ACS as the Nation's preeminent source of sub-national income and poverty estimates. At this point we really don't come close to understanding the reasons why income estimates from Census 2000 were quite a bit higher than the C2SS, though we have thoughts on some of the factors behind these differences. More work has to be done on examining which of these factors played major roles in these differences. One of the Census 2000 evaluation projects, the CPS/Census 2000 Exact Match Project, may shed light on Census 2000 income reporting/imputation patterns that may in turn shed light on why the C2SS and Census 2000 figures differed. It is clear that we are just at the beginning stages of understanding why Census 2000 and C2SS income figures differ.

REFERENCE

Posey, Kirby G., and Edward Welniak, *Income in the ACS: Comparisons to the 1990 Census*, U.S. Census Bureau, Washington, DC 20233, March 1998, Revised June 1999.

Table 1: Comparison of Median Income of Individuals by Income Type and Reference Period (asked October - December, 1997)

ACS SPLIT PANEL TEST

	PAST 12 MONTHS	CALENDAR YEAR 1996	DIFFERENCE	
Wages or Salary	$25,000	$26,000	-$1,000	*
Self-Employment	$12,000	$10,000	$2,000	*
Interest	$1,500	$1,400	$100	
Social Security	$7,932	$8,028	-$96	
Retirement	$9,700	$10,000	-$300	
Public Assistance	$3,438	$3,872	-$434	
Other Income	$3,000	$3,000	$0	
Total Income	$24,000	$24,662	-$662	

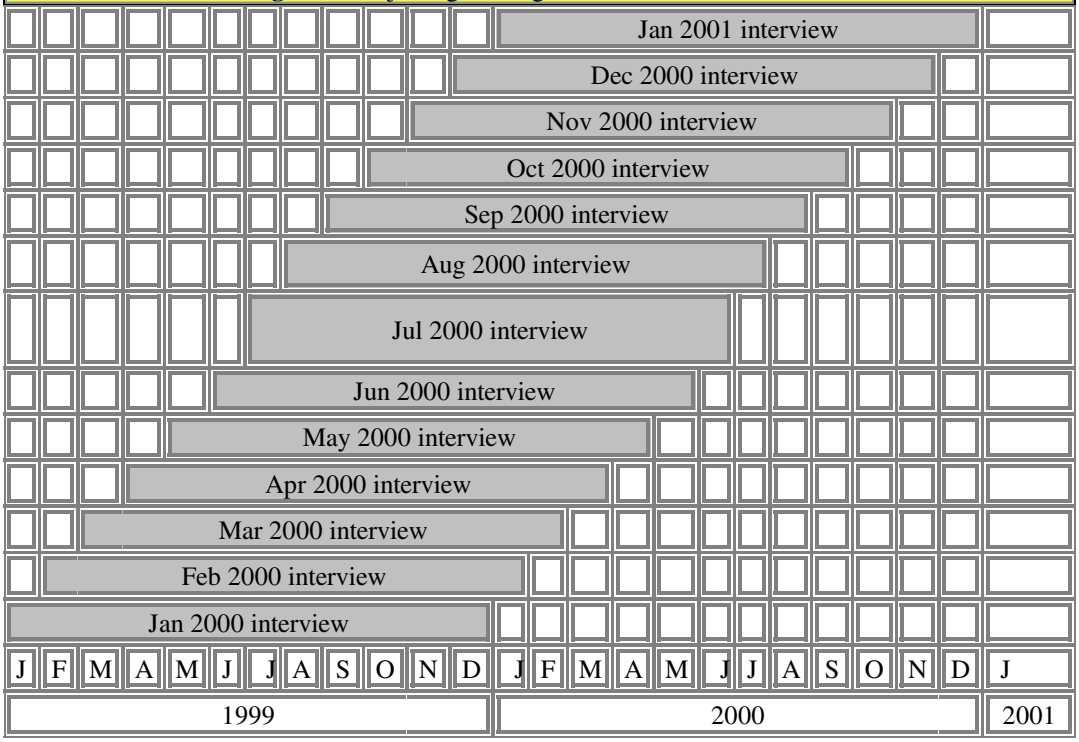
Note: These results are from mail responses only.
Source: U.S. Census Bureau, 1996 American Community Survey, Split Panel Test
* Indicates statistically significant difference at 90-percent confidence level.

Table 2 - Response Rates for All Sources of Income - Split Panel Test

Income	Period	Percent Reporting Income	Difference	
Wage/Salary	Past 12 months	73.93%		
	Calendar year 1996	72.84%	+1.09	
Self-Employment	Past 12 months	73.72%		
	Calendar year 1996	72.68%	+1.04	
Interest, etc.	Past 12 months	72.07%		
	Calendar year 1996	71.05%	+1.02	
Social Security	Past 12 months	75.78%		
	Calendar year 1996	74.65%	+1.13	
Retirement	Past 12 months	75.36%		
	Calendar year 1996	73.96%	+1.40	
Public Assistance	Past 12 months	77.12%		
	Calendar year 1996	75.38%	+1.74	*
Other Income	Past 12 months	76.40%		
	Calendar year 1996	75.23%	+1.17	
Total Income	Past 12 months	83.67%		
	Calendar year 1996	83.09%	+0.58	

Source: 1996 American Community Survey, Split Panel Test
* Indicates statistically significant difference at 90-percent confidence level.

Figure 1: Adjusting Rolling Reference Period Income:



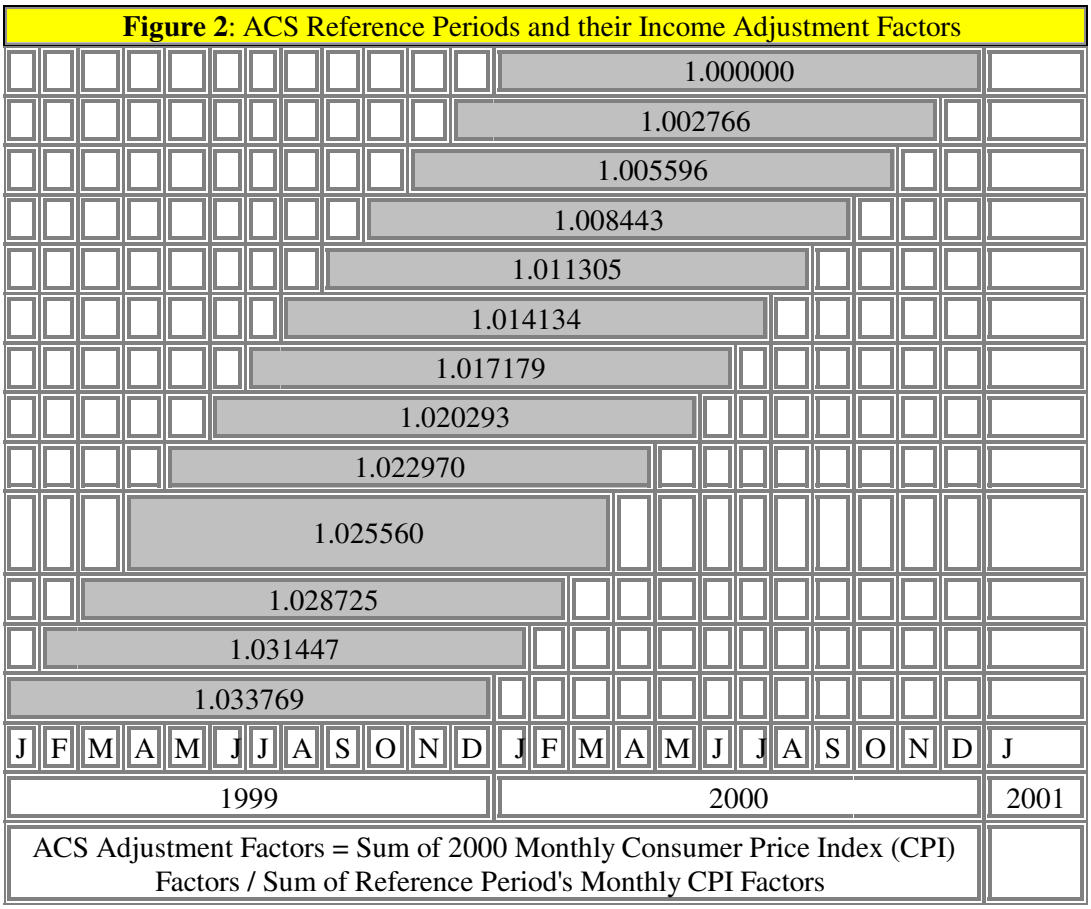


Table 3. Comparisons of Median Household Income in 2000 Dollars: C2SS/ACS, Census 2000, and the CPS

State	C2SS/ ACS Estimate ($)	Census 2000 ($)	Percent Difference (ACS-Census)	Statistically Significant Difference (denoted by "*")	Current Population Survey (CPS) (4)	Percent Difference (ACS-CPS)	Statistically Significant Difference (denoted by "*")
United States	41,486	43,396	-4.6	*	41,990	-1.2	*
Alabama	33,433	35,274	-5.5	*	35,424	-6.0	*
Alaska	52,906	53,292	-0.7		52,847	0.1	
Arizona	38,547	41,912	-8.7	*	39,783	-3.2	
Arkansas	33,188	33,256	-0.2		29,697	10.5	*
California	46,617	49,078	-5.3	*	46,816	-0.4	
Colorado	46,391	48,778	-5.1		48,240	-4.0	
Connecticut	54,129	55,735	-3.0	*	50,172	7.3	*
Delaware	47,695	48,962	-2.7		50,365	-5.6	
District of Columbia	40,926	41,466	-1.3		41,222	-0.7	
Florida	38,054	40,115	-5.4	*	38,856	-2.1	
Georgia	41,295	43,849	-6.2	*	41,901	-1.5	
Hawaii	51,587	51,483	0.2		51,546	0.1	
Idaho	37,569	38,826	-3.3		37,611	-0.1	

State	C2SS/ ACS Estimate ($)	Census 2000 ($)	Percent Difference (ACS-Census)	Statistically Significant Difference (denoted by "*")	Current Population Survey (CPS) (4)	Percent Difference (ACS-CPS)	Statistically Significant Difference (denoted by "*")
Illinois	45,708	48,145	-5.3	*	46,064	-0.8	
Indiana	40,794	42,954	-5.3	*	40,865	-0.2	
Iowa	38,179	40,786	-6.8	*	40,991	-7.4	*
Kansas	40,264	41,980	-4.3	*	41,059	-2.0	
Kentucky	32,862	34,796	-5.9	*	36,265	-10.4	*
Louisiana	30,985	33,653	-8.6	*	30,718	0.9	
Maine	36,608	38,483	-5.1	*	37,266	-1.8	
Maryland	52,447	54,632	-4.2	*	54,535	-4.0	
Massachusetts	49,801	52,188	-4.8	*	46,753	6.1	*
Michigan	43,387	46,158	-6.4	*	45,512	-4.9	*
Minnesota	47,753	48,683	-1.9		54,251	-13.6	*
Mississippi	32,728	32,376	1.1		34,299	-4.8	
Missouri	37,145	39,200	-5.5	*	45,097	-21.4	*
Montana	33,103	34,126	-3.1		32,777	1.0	
Nebraska	37,379	40,560	-8.5	*	41,750	-11.7	*
Nevada	42,401	46,069	-8.7	*	45,758	-7.9	*
New Hampshire	50,969	51,118	-0.3		50,926	0.1	
New Jersey	54,276	56,986	-5.0	*	50,405	7.1	*
New Mexico	33,428	35,272	-5.5	*	35,093	-5.0	
New York	43,735	44,841	-2.5	*	40,744	6.8	*
North Carolina	37,784	40,492	-7.2	*	38,317	-1.4	
North Dakota	34,376	35,759	-4.0		35,996	-4.7	
Ohio	39,745	42,323	-6.5	*	42,962	-8.1	*
Oklahoma	34,135	34,515	-1.1		32,432	5.0	*
Oregon	39,090	42,282	-8.2	*	42,499	-8.7	*
Pennsylvania	39,661	41,445	-4.5	*	42,176	-6.3	*
Rhode Island	43,778	43,495	0.6		42,197	3.6	
South Carolina	36,439	38,320	-5.2	*	37,570	-3.1	
South Dakota	34,840	36,460	-4.6	*	36,475	-4.7	*
Tennessee	36,559	37,574	-2.8	*	34,096	6.7	*
Texas	39,398	41,260	-4.7	*	38,609	2.0	
Utah	45,536	47,252	-3.8		47,550	-4.4	
Vermont	40,505	42,220	-4.2	*	39,594	2.2	
Virginia	47,125	48,235	-2.4	*	47,163	-0.1	
Washington	45,246	47,304	-4.5	*	42,525	6.0	*
West Virginia	29,089	30,687	-5.5	*	29,411	-1.1	
Wisconsin	42,209	45,253	-7.2	*	45,088	-6.8	*
Wyoming	38,614	39,157	-1.4		39,629	-2.6	