



Los Angeles American Indian and Alaska Native Project¹

Technical Memo 5: AIAN Underrepresentation in the ACS

Jonathan Ong and Paul Ong
November 19, 2012

This technical memo addresses an underestimation of American Indians and Alaska Natives (AIAN) in the American Community Survey (ACS) relative to the number reported in the 2010 decennial enumeration. This problem is due largely to a severe underestimation of the single-race AIAN population, which is comprised of those who self-identify as being AIAN alone.

The decennial count is the official count performed by the U.S. Census Bureau and is conducted to get as complete a count of the population of the United States as feasible. The enumeration is conducted in accordance with Article I, Section 2 of the US Constitution, and the data are used for reapportionment (the allocation of Congressional seats among states) and redistricting (the drawing of electoral districts). Race data are collected in part to help enforce voting rights for minorities (U.S. Census Bureau, n.d). The enumeration collects information on the age, race, ethnicity, gender, household size, and household type. Although the decennial count suffers from an undercount of some disadvantaged groups (U.S. Census Bureau, 2012), the unadjusted numbers are used by the federal government as official figures. For the purpose of this memo, the counts are taken as the baseline.

The ACS has replaced the census long-form survey, which was conducted at the same time of the decennial enumeration. The ACS is conducted continuously and reported annually, and collects demographic,

socioeconomic, and housing data. The results are widely used by private, public, and nonprofit agencies and organizations. Any serious flaws in ACS statistics can have serious implications in identifying problems, shaping policies, and allocating resources. Unlike the decennial enumeration, the ACS is based on a small sample that is roughly 2.5% of all households in the United States each year. The sample is weighted to generate estimates of the total population. The accuracy of those estimates depends on the representativeness of the sample and the precision of the weights. Moreover, the ACS is subject to sampling error, and the “luck of the draw” may randomly generate statistical discrepancies.

For this memo, we examine the 2009–11 ACS in order to have a significant number of AIANs in the ACS sample. Both the decennial enumeration and the ACS are based on self-reported racial identity; consequently, a respondent can report being AIAN alone or AIAN in combination with another racial category. AIANs are classified three ways: (1) an inclusive count of those who are AIAN alone or in combination; (2) a single-race or alone count of those who selected only the AIAN racial category; and (3) a mixed-race or in-combination count of those who self-reported being AIAN and some other race.

As seen in table 1, the ACS estimates for the total population closely mirror the total population in the 2010 enumeration, which is not surprising because the weights are based on annual population estimations, which are benchmarked to the 2010 Decennial Enumeration. The 2009–11 ACS estimate of the total population count is within a fraction of a percentage point of the 2010 decennial census. The difference may be due to nonlinear growth between 2009 and 2011 and the fact that the decennial enumeration was conducted in April of 2010 rather than the middle of 2010.

¹This technical memo is a product of a collaborative effort by UCLA American Indian Studies Center and the Los Angeles Urban Indian Roundtable. We would like to thank reviewers for their input, feedback, and comments. The authors are solely responsible for the contents of this report.

Table 1: 2010 Enumeration Counts and ACS Estimates

	Total	AIAN Alone or in Combination	AIAN Alone Only	AIAN in Combination Only
2010 Decennial Enumeration				
United States	308745538	5220579	2932248	2288331
California	37253956	723225	362801	360424
Los Angeles	9818605	140764	72828	67936
2011 ACS 3-year Average				
United States	309231244	5055427	2529104	2526323
California	37330448	677223	287184	390039
Los Angeles	9834410	132488	46946	85542
ACS-to-Decennial Ratio				
United States	100.2%	96.8%	86.2%	110.4%
California	100.2%	93.6%	79.2%	108.2%
Los Angeles	100.2%	94.1%	64.5%	125.9%

While there is virtually no discrepancy for the total population, the inclusive AIAN estimates are noticeably lower, as reported in Table 1. For the inclusive AIAN counts (alone and in combination), the ACS national estimate is about 3% lower than the 2010 counts. The discrepancy is greater for California and Los Angeles² (about 4% lower). The difference is due entirely to an underestimation of AIANs alone. At the national level, the ACS estimates are less than seven-eighths of the 2010 counts. This is partially offset by a higher estimate by the ACS of AIANs in combination, which is about a tenth greater than the decennial numbers. This pattern of over- and underestimation is repeated for California. There are only four AIANs alone in the ACS for every five in the 2010 census, partially offset by higher figures in the ACS for AIANs in combination. The problem is most severe for Los Angeles, with the 2009–11 ACS reporting less than two-thirds of the AIAN alone than the 2010 census.³

The discrepancies between ACS and the decennial counts cannot be explained by random sampling error. The difference is much larger than the margin of error for the ACS. For example, the margin of error for AIAN alone in Los Angeles County is +/-2,603; however, the observed discrepancy is nearly 26,000 fewer AIAN alone in the ACS

tabulations than in the 2010 count, a difference of nearly tenfold the margin of error. This is not unique to Los Angeles alone and thus appears to be a systemic problem. In California the margin of error given is +/-7,627, but the difference is greater than 75,000, resulting again in a nearly tenfold difference. On the national level the margin of error given is +/-17,900, but the difference is about 403,000, which is more than a twentyfold difference.

There are two potential sources for the underestimation in the ACS: under sampling due to a difficulty in reaching AIANs and incorrect weights used to convert the sample into estimates. One possible way to determine the relative importance of the two possible sources is to analyze the 2009–11 ACS Public Use Micro-Sample (PUMS) data. PUMS is an individual-level subsample of ACS data, covering approximately 1% of the population. Tabulating the ACS PUMS produces very similar population estimates as the web-published ACS tables. Table 2 reports the weighted estimates of the AIAN populations as a proportion of the total population, and most of the percentages are within 0.01 points. Because of the similarity, analyzing ACS PUMS can provide reasonable insights into the source of the underestimation problem.

²Los Angeles refers to Los Angeles County, which is the same geography used for the Los Angeles Metropolitan Statistical Area (MSA). Note that this is not the same MSA as the Los Angeles-Anaheim MSA which includes more than just Los Angeles County.

³We also find the same pattern of underestimation and overestimation for the vast majority of metropolitan areas with large AIAN populations. See appendix.

Table 2: ACS Public and PUMS Estimates of AIAN as % of Total Population

	AIAN Alone or in Combination	AIAN Alone Only	AIAN in Combination Only
United States			
2009–11 ACS Public Estimates	1.63%	0.82%	0.82%
2009–11 ACS PUMS Weighted	1.62%	0.81%	0.81%
California			
2009–11 ACS Public Estimates	1.81%	0.77%	1.04%
2009–11 ACS PUMS Weighted	1.81%	0.77%	1.04%
Los Angeles			
2009–11 ACS Public Estimates	1.35%	0.48%	0.87%
2009–11 ACS PUMS Weighted	1.37%	0.49%	0.88%

Table 3 reports AIANs as a percent of the unweighted PUMS sample. The three-year PUMS sample contains all AIANs (alone and in combination), roughly proportionate to the AIAN share of the decennial counts and in fact slightly higher. Over sampling is also noticeable for AIANs in combination at all-three geographic levels. AIANs alone, however, are under-sampled in California and more so for

Los Angeles. In this region, they made up more than 0.7% of people counted in the 2010 decennial census but less than 0.6% of the ACS sample. While sampling variation contributes to the inaccurate ACS estimates of the AIAN populations, systematic sampling error cannot explain the entire discrepancy.

Table 3: AIANs as a Percent of 2010 Count and Unweighted ACS Sample

	AIAN Alone or in Combination	AIAN Alone Only	AIAN in Combination Only
United States			
2010 Decennial Enumeration	1.69%	0.95%	0.74%
2009–11 ACS PUMS Unweighted	1.83%	1.00%	0.83%
California			
2010 Decennial Enumeration	1.94%	0.97%	0.97%
2009–11 ACS PUMS Unweighted	2.00%	0.91%	1.09%
Los Angeles			
2010 Decennial Enumeration	1.43%	0.74%	0.69%
2009–11 ACS PUMS Unweighted	1.47%	0.59%	0.88%

The problem is compounded by systematic differences in the weights used to translate the ACS sample into population estimates. This can be seen in table 4. The mean weight across all three years is slightly over 100, corresponding to the approximate 1% sampling, but the mean weight for AIAN alone is only 84, 87, and 82 in Los Angeles, California, and the nation respectively. This is reasonable when a group is over-sampled, such as the case

for AIANs nationally, although the weights are still too low. In Los Angeles, the lower AIAN-alone weights interact with under-sampling to produce the severe underestimate of AIANs alone in the ACS.

Table 4: Mean Weights in 2009–11 Weights

	Total Population	AIAN Alone or in Combination	AIAN Alone Only	AIAN in Combination Only
United States	100.8	89.2	82.0	97.9
California	103.5	93.7	87.2	99.2
Los Angeles	102.1	94.7	83.9	101.9

The above findings have potentially profound implications for our understanding of the AIAN population, the problems they face, and the policies based on government statistics. According to the U.S. Census Bureau:

The American Community Survey (ACS) is an ongoing survey that provides data every year—giving communities the current information they need to plan investments and services. Information from the survey generates data that help determine how more than \$400 billion in federal and state funds are distributed each year.

http://www.census.gov/acs/www/about_the_survey/resources/congress.php

Inaccurate reporting of the size of the AIAN population can place this group at a disadvantage, and this is particularly true for AIANs alone. This can lead to an undercount of the number of AIANs facing various employment, educational, and other challenges. The analysis indicates that both sampling and weighting can contribute to the discrepancy between ACS and the decennial count. The relative roles of these two factors vary geographically.⁴

Based on the findings, we recommend that the Census Bureau reweigh the data to improve the population estimates. There are, however, other issues, such as the severe under sampling of AIANs alone in Los Angeles. This will require additional research to identify the cause of this problem, research that should be supported by the Census Bureau in consultation with AIAN experts.

REFERENCES

“Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census.” *Newsroom: 2010 Census: Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census*. U.S. Census Bureau, May 22 2012. Web. Aug. 24 2012. <http://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html>.

“Census in the Constitution.” *Census in the Constitution—2010 Census*. U.S. Census Bureau, n.d. Web. Aug. 24 2012. <<http://2010.census.gov/2010census/about/constitutional.php>>.

We would like to thank our sponsors, The California Wellness Foundation, Los Angeles County Board of Supervisor Don Knabe, and the UCLA Center for the Study of Inequality, for their generous support. We would also like to thank the authors, Paul Ong and Jonathan Ong, as well as the American Indian Studies Center for supporting this project.

⁴Another potential source is differences in the way individuals respond to the race question in ACS and the enumeration, but that is impossible to test with existing data. The questions, however, are identical.

APPENDIX

Geographies	2010 Decennial Enumeration			2011 ACS 3-year Average		
	Inclusive	Alone	Combination	Inclusive	Alone	Combination
<i>United States</i>	1.69%	0.95%	0.74%	1.63%	0.82%	0.82%
<i>California</i>	1.94%	0.97%	0.97%	1.81%	0.77%	1.04%
California MSAs						
Los Angeles-Long Beach-Santa Ana	1.39%	0.71%	0.68%	1.25%	0.46%	0.79%
San Francisco-Oakland-Fremont	1.53%	0.57%	0.96%	1.40%	0.44%	0.97%
Riverside-San Bernardino-Ontario	2.07%	1.10%	0.97%	2.02%	0.99%	1.03%
San Diego-Carlsbad-San Marcos	1.70%	0.85%	0.85%	1.75%	0.63%	1.11%
Sacramento--Arden-Arcade--Roseville	2.49%	1.01%	1.49%	2.52%	0.99%	1.52%
San Jose-Sunnyvale-Santa Clara	1.53%	0.75%	0.78%	1.33%	0.57%	0.76%
Fresno	2.74%	1.68%	1.06%	2.16%	0.95%	1.21%
Bakersfield-Delano	2.69%	1.51%	1.18%	2.42%	1.19%	1.23%
Oxnard-Thousand Oaks-Ventura	1.94%	0.98%	0.96%	1.87%	0.82%	1.05%
Stockton	2.40%	1.05%	1.35%	2.63%	1.19%	1.44%
Modesto	2.46%	1.15%	1.31%	2.31%	0.99%	1.32%
Santa Rosa-Petaluma	2.75%	1.34%	1.41%	2.43%	1.34%	1.09%
Visalia-Porterville	2.61%	1.58%	1.02%	2.54%	1.29%	1.26%
Santa Barbara-Santa Maria-Goleta	2.39%	1.29%	1.09%	1.96%	1.07%	0.90%
Salinas	2.31%	1.32%	1.00%	2.18%	1.19%	0.99%
Vallejo-Fairfield	2.27%	0.78%	1.50%	1.94%	0.62%	1.32%
San Luis Obispo-Paso Robles	2.18%	0.94%	1.24%	1.97%	0.86%	1.11%
Santa Cruz-Watsonville	2.10%	0.86%	1.24%	1.69%	0.46%	1.24%
Merced	2.32%	1.36%	0.96%	1.61%	0.89%	0.73%
Chico	4.12%	2.00%	2.12%	4.24%	1.14%	3.10%
Redding	5.29%	2.79%	2.50%	4.79%	2.29%	2.50%
El Centro	2.37%	1.75%	0.62%	3.29%	1.91%	1.38%
Hanford-Corcoran	2.60%	1.67%	0.92%	2.29%	1.29%	1.00%
Madera-Chowchilla	4.08%	2.74%	1.34%	3.36%	1.83%	1.54%
Napa	1.76%	0.78%	0.98%	1.74%	0.79%	0.95%
Other MSAs						
New York-Northern New Jersey-Long Island, NY-NJ-PA	1.03%	0.49%	0.54%	0.73%	0.27%	0.46%
Chicago-Joliet-Naperville, IL-IN-WI	0.83%	0.39%	0.44%	0.64%	0.23%	0.41%
Dallas-Fort Worth-Arlington, TX	1.32%	0.68%	0.64%	1.39%	0.53%	0.86%
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.81%	0.27%	0.53%	0.67%	0.17%	0.50%
Houston-Sugar Land-Baytown, TX	1.17%	0.64%	0.52%	0.95%	0.42%	0.53%
Washington-Arlington-Alexandria, DC-VA-MD-WV	1.11%	0.41%	0.70%	0.99%	0.37%	0.62%
Phoenix-Mesa-Glendale, AZ	3.16%	2.37%	0.79%	2.82%	2.15%	0.68%
Seattle-Tacoma-Bellevue, WA	2.43%	1.07%	1.36%	2.35%	0.98%	1.38%
Oklahoma City, OK	7.18%	4.09%	3.08%	7.90%	3.63%	4.28%
Tulsa, OK	13.17%	8.25%	4.91%	13.04%	6.89%	6.15%
Albuquerque, NM	7.16%	5.86%	1.30%	6.97%	5.78%	1.19%
Farmington, NM	38.74%	36.63%	2.11%	37.68%	36.61%	1.07%

Geographies	ACS-to-Decennial Ratio			
	Total	Inclusive	Alone	Combination
<i>United States</i>	100.2%	96.8%	86.3%	110.4%
<i>California</i>	100.2%	93.6%	79.2%	108.2%
California MSAs				
Los Angeles-Long Beach-Santa Ana	100.2%	90.1%	64.9%	116.3%
San Francisco-Oakland-Fremont	100.2%	91.6%	76.4%	100.7%
Riverside-San Bernardino-Ontario	100.3%	98.0%	90.7%	106.2%
San Diego-Carlsbad-San Marcos	100.2%	102.8%	74.7%	130.9%
Sacramento--Arden-Arcade--Roseville	100.2%	101.1%	99.1%	102.5%
San Jose-Sunnyvale-Santa Clara	100.3%	87.1%	76.4%	97.4%
Fresno	100.2%	79.0%	56.7%	114.6%
Bakersfield-Delano	100.2%	90.0%	78.9%	104.2%
Oxnard-Thousand Oaks-Ventura	100.1%	96.8%	84.2%	109.7%
Stockton	100.3%	110.1%	113.7%	107.4%
Modesto	100.1%	94.1%	86.2%	101.0%
Santa Rosa-Petaluma	100.0%	88.4%	99.8%	77.4%
Visalia-Porterville	100.2%	97.8%	81.5%	123.1%
Santa Barbara-Santa Maria-Goleta	100.0%	82.4%	82.4%	82.4%
Salinas	100.3%	94.7%	90.7%	99.9%
Vallejo-Fairfield	100.1%	85.3%	79.6%	88.2%
San Luis Obispo-Paso Robles	100.2%	90.6%	91.8%	89.6%
Santa Cruz-Watsonville	100.0%	80.6%	53.2%	99.6%
Merced	100.2%	69.6%	65.4%	75.4%
Chico	100.0%	102.8%	57.0%	145.8%
Redding	100.1%	90.6%	82.1%	100.1%
El Centro	100.1%	139.0%	108.8%	224.5%
Hanford-Corcoran	100.1%	88.0%	76.8%	108.2%
Madera-Chowchilla	100.2%	82.7%	66.7%	115.3%
Napa	100.2%	99.0%	101.9%	96.7%
Other MSAs				
New York-Northern New Jersey-Long Island, NY-NJ-PA	100.1%	71.6%	55.6%	86.3%
Chicago-Joliet-Naperville, IL-IN-WI	100.1%	77.0%	58.5%	93.1%
Dallas-Fort Worth-Arlington, TX	100.5%	105.9%	78.3%	135.2%
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	100.1%	82.8%	61.9%	93.5%
Houston-Sugar Land-Baytown, TX	100.4%	81.8%	66.0%	101.2%
Washington-Arlington-Alexandria, DC-VA-MD-WV	100.4%	89.2%	91.5%	87.9%
Phoenix-Mesa-Glendale, AZ	100.4%	89.7%	91.0%	85.7%
Seattle-Tacoma-Bellevue, WA	100.4%	97.1%	91.5%	101.6%
Oklahoma City, OK	100.4%	110.6%	89.0%	139.4%
Tulsa, OK	100.2%	99.2%	83.6%	125.3%
Albuquerque, NM	100.1%	97.4%	98.8%	91.3%
Farmington, NM	99.4%	96.7%	99.3%	50.4%