



California and the 2000 Census

Introduction

As the U.S. Census Bureau prepares to conduct the nation's next decennial census on April 1, 2000, controversy exists concerning plans to supplement the traditional population headcount with estimates derived from statistical sampling techniques.

LAO Findings

In 1990, the nation's population was undercounted by roughly 1.6 percent. California's undercount, however, was much higher—2.7 percent. This higher undercount likely cost California one seat in the U.S. House of Representatives and at least \$2 billion in federal funds during the 1990s.

Different subgroups of the population have been undercounted at different rates. For example, undercount rates were highest for minorities, renters, and those persons living in nontraditional households. As a result, the undercount rates varied significantly among counties. For instance, in 1990 Los Angeles County and most counties in the San Joaquin Valley had rates much higher than the statewide average.

The bureau currently plans to release two sets of population figures—one using traditional counting techniques and a second set which uses sampling to correct for the undercount. Congress, however, has not as yet provided funding for the sampling process.

If sampling-adjusted numbers are provided, state policymakers will have to decide which set of data to use for redistricting in California and for allocating state funds to localities.



BACKGROUND

In about a year, the first decennial census of the new millennium—Census 2000—will be conducted. The decennial census is the single most important source of information about the people of the United States, and the forthcoming census will be the twenty-second that has taken place over the past 200 years, beginning in 1790.

Since that first census, each decennial census has attempted to count each and every person in the country via direct contact. With recent censuses, each household receives a questionnaire to answer and return via the U.S. mail. Households that do not respond to the questionnaire are subsequently visited by census staff. This self-enumeration approach is rooted in the concept of relying on a minimally intrusive process and respect for individual privacy.

In contrast to this method, some governments use more invasive approaches. In Turkey, for example, its latest census involved counting the entire population manually in one day over a 14-hour period, with people being required to stay home and be counted under threat of punishment if found in public without special permission during this time interval.

Why the Census Is So Important

The decennial census is important because it provides the only true statistical “snapshot” of the entire U.S. population—both in terms of its size and characteristics. It is used for a wide variety of purposes—by economists and the business com-

munity in documenting demographic trends and their implications, by policymakers to understand the characteristics of the population and its needs, by governments to allocate spending to different governmental entities, and by federal authorities to determine the allocation/apportionment of electoral districts and how many representatives each state will have in the Congress. Given these uses, the census is an extremely important undertaking, and its integrity and accuracy are of paramount importance.

Current Controversies Regarding the Census

Conducting the census is inherently an extremely challenging undertaking involving thousands of census workers, a budget in the billions of dollars, and requiring the voluntary cooperation of hundreds of millions of people. The challenge is complicated by the sheer physical size of the country, its geographic diversity, the mobility of its population, its high rates of foreign in-migration, and its nonhomogeneous population reflecting its great ethnic diversity and wide mix of urban, suburban, and rural communities. It also appears that Americans are becoming a bit less responsive to the census questionnaire process. Given this, conducting the census today is a truly massive and difficult undertaking, and inherently open to debate and disagreement about how it can best be carried out and/or improved.

History indicates that problems with obtaining accurate census counts are not new. Even when

Thomas Jefferson—who headed-up the first census—reported the results, he noted that there was evidence that some persons had been missed. Over the years, as the country has continued to expand and society has undergone changes, there always have been new challenges to obtaining accurate census counts.

As the U.S. Census Bureau approaches the 2000 census, however, an unusual degree of controversy exists. The key issue is: Should a traditional headcount be relied on as in past years, or should this headcount be supplemented with statistical estimates to account for persons missed by the enumerators? This has been referred to as the statistical sampling debate, and has filled the newspaper headlines on-and-off, been the topic of conferences and symposiums for economists and demographers, resulted in Congressional hearings, and been the subject of litigation.

Why Was Sampling Proposed?

The interest in statistical sampling evolved in response to documented problems experienced in previous decennial censuses in accurately measuring the population. The U.S. Census Bureau had previously determined that its decennial censuses were resulting in population *undercounts*, and the idea of sampling was proposed in part to deal with them. The fact that particularly significant undercounts occurred in the 1990 census has been especially well documented and publicized in recent years.

Of particular concern has been the fact that states experiencing larger-than-average undercounts have been “shorted” federal funds, given that many federal funding formulas use population as a factor in determining the share of funds going to individual states. As discussed below, California is significantly affected by undercounting.

WHAT IS KNOWN ABOUT THE 1990 NATIONAL UNDERCOUNT?

The Causes and Magnitude of the Undercount

Census undercounts can occur for two basic reasons: (1) the “master list” of households used to identify people is not completely accurate, and (2) difficulties are encountered in tracking down and collecting information on individuals who live in households. Figure 1 (see page 4) provides the Census Bureau’s own estimates of the census

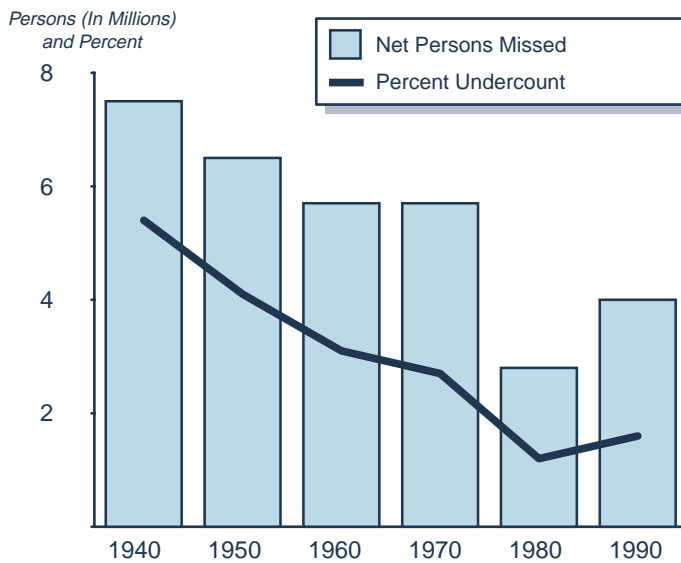
undercount since 1940. While the accuracy of the census improved steadily (that is, the undercount declined) between 1940 and 1980, the 1990 census took a step backward on the fundamental issue of accuracy. Indeed, the 1990 count missed 4 million people, an error margin of 1.6 percent. More troubling, the 1990 census was the first to be less accurate than its predecessor.



Figure 1

National Census Undercount

1940 Through 1990



The Undercount Differs By Population Subgroup

Compounding the undercounting problem is that different subgroups of the population are undercounted to different degrees. For instance, African Americans tend to be undercounted to a greater degree than the population generally. After the 1940 census, for example, the Census Bureau gave the Selective Service an estimate of how many young men it could expect to answer its call for the war effort. In total, 3 percent more men registered for the draft than had been counted by enumerators. Among the African-American community, however, 13 percent more men showed up for registration than had been expected based on 1940 census data. Similarly, in the 1990

census, it has been estimated that African Americans had a net undercount of 4.4 percent, compared to only 1.2 percent for non-African Americans. The latter included 0.9 percent for whites and 2.3 percent for Asian Americans.

Factors Contributing to the Undercounting Problem

Undercounting is caused by many factors:

- ◆ **Increased Population Mobility.** During the period 1990 to 1994, 17 percent of the American population on average changed residences each year.

This increased mobility makes locating households harder for census-count purposes.

- ◆ **Changing Domestic Living Arrangements.** Households have always been the major focus of census enumeration. In eighteenth century America, nearly all citizens identified themselves with a household whose members were almost always related by blood, marriage, or through regular employment and, therefore, included servants, apprentices, and resident farmworkers. Most people lived in a family-occupied dwelling, and it was much easier to provide a population count and characterize the members of the house-

ESTIMATING THE 1990 CENSUS UNDERCOUNT

Prior to 1990, the Census Bureau primarily relied on “demographic analysis” to estimate the level of the undercount. This analysis relies on such factors as administrative records of births, deaths, immigration, and emigration to provide estimates of the true population total.

In 1990, the Census Bureau used an additional technique to generate estimates of the undercount—the Post Enumeration Survey (PES). Essentially, once the 1990 census was conducted, the bureau drew a sample of census blocks from around the country. Then, census staff knocked on the door of each housing unit in the sample census blocks—regardless of whether it was on the master address list. For each of the blocks, the Census Bureau compared the information from both sources—that is, the official census and the PES. Based on the results of the PES survey, an estimate was made of people missed in the original census count. Then, using a statistical model, these PES estimates of undercount were used to develop undercount figures for all geographic areas.

Census research indicates that the majority of the undercount is associated with incorrect reporting during the initial census. The PES relies on an intensive effort (much more so than the initial census) to count all the residents in the sample blocks. This “saturation coverage” is the key to the survey’s ability to identify the undercount.

hold. Today, divorce, cohabitation without marriage, and group housing make the determination of whom to count and where to count them increasingly complex. From 1970 through 1990 alone, the number of American households grew 47 percent, while average household size shrank from 3.1 persons to 2.6 persons, and nonfamily households grew by 128 percent. Moreover, because higher proportions of the nation’s children, renters, and minorities experience these living arrangements, this contributes to

their undercount rates being higher than for the population generally.

- ◆ **Other Factors.** Other factors which influence undercounting include language barriers inhibiting the reading of census forms and responding to interviews; habitation in irregular housing units, such as illegal units, certain mobile homes, and secured buildings; and neighborhood conditions that lead to resistance to outsiders, concealment to protect resources, and disbelief of census confidentiality.



Reduced Accuracy Occurred Despite Increased Funding

The 1990 census was the most expensive in history, costing \$25 per housing unit. In contrast, on an inflation-adjusted basis, the 1970 census cost only \$11 per housing unit, and the 1980 census cost \$20 per housing unit. Therefore, even after accounting for inflation and increases to population, the 1990 census cost twice as much as the 1970 census.

Much of this cost increase can be explained by the decline in the percentage of households that returned the census questionnaire by mail, and the resulting need of more expensive, labor-intensive follow-up procedures using hundreds of census takers going door-to-door. When census questionnaires were mailed in 1970, 78 percent of housing units mailed back their questionnaires. By 1990, that percentage had fallen to 65 percent.

Thus, it has been argued that the 1990 census failed on two fronts: (1) it was too expensive and (2) it counted too few people. It is because of this experience of the 1990 census that many economic and demographic experts hold the view that it has become both physically impossible and cost-prohibitive to even attempt to directly count every person in the United States.

How Big was California's 1990 Undercount?

The 1990 census undercount varied significantly by state. In large part this reflected the fact discussed above that undercount rates tend to differ for different population characteristics and living situations, and these are not the same in all states. In addition, the extent to which a state's population lives in urban versus suburban versus rural settings is a factor.

As noted above, the net 1990 national undercount was estimated to be 4 million people, or 1.6 percent of the population. As shown in Figure 2, California's undercount was disproportionately worse—an estimated 835,000, or 2.7 percent, of the state's population, was missed. In terms of the *number* of people undercounted, its 835,000 undercount was almost double that of Texas, the state with the second-highest numerical undercount. In *percentage* terms, the state's undercount rate was fourth highest—trailing only the District of Columbia, New Mexico, and Texas. The geographic distribution of the state's undercount is discussed later.

Figure 2

1990 Census Undercount, by Region and State

Region/State	1990 Population Total		Estimated Undercount	
	Reported	Adjusted for Estimated Undercount	Amount	Percent
Northeast:				
Connecticut	3,287,116	3,308,343	21,227	0.6%
Maine	1,227,928	1,237,130	9,202	0.7
Massachusetts	6,016,425	6,045,224	28,799	0.5
New Hampshire	1,109,252	1,118,632	9,380	0.8
New Jersey	7,730,188	7,774,461	44,273	0.6
New York	17,990,455	18,262,491	272,036	1.5
Rhode Island	1,003,464	1,004,815	1,351	0.1
Pennsylvania	11,881,643	11,916,783	35,140	0.3
Vermont	562,758	569,100	6,342	1.1
Midwest:				
Illinois	11,430,602	11,544,319	113,717	1.0%
Indiana	5,544,159	5,572,057	27,898	0.5
Iowa	2,776,755	2,788,332	11,577	0.4
Kansas	2,477,574	2,495,014	17,440	0.7
Michigan	9,295,297	9,361,308	66,011	0.7
Minnesota	4,375,099	4,394,610	19,511	0.4
Missouri	5,117,073	5,148,974	31,901	0.6
Nebraska	1,578,385	1,588,712	10,327	0.7
North Dakota	638,800	643,033	4,233	0.7
Ohio	10,847,115	10,921,741	74,626	0.7
South Dakota	696,004	702,864	6,860	1.0
Wisconsin	4,891,769	4,921,871	30,102	0.6

Continued



Region/State	1990 Population Total		Estimated Undercount	
	Reported	Adjusted for Estimated Undercount	Amount	Percent
South:				
Alabama	4,040,587	4,113,810	73,223	1.8%
Arkansas	2,350,725	2,392,596	41,871	1.8
Delaware	666,168	678,385	12,217	1.8
District of Columbia	606,900	628,309	21,409	3.4
Florida	12,937,926	13,197,755	259,829	2.0
Georgia	6,478,216	6,620,641	142,425	2.2
Kentucky	3,685,296	3,746,044	60,748	1.6
Louisiana	4,219,973	4,314,085	94,112	2.2
Maryland	4,781,468	4,882,452	100,984	2.1
Mississippi	2,573,216	2,629,548	56,332	2.1
North Carolina	6,628,637	6,754,567	125,930	1.9
Oklahoma	3,145,585	3,202,963	57,378	1.8
South Carolina	3,486,703	3,559,547	72,844	2.0
Tennessee	4,877,185	4,964,261	87,076	1.8
Texas	16,986,510	17,472,538	486,028	2.8
Virginia	6,187,358	6,313,836	126,478	2.0
West Virginia	1,793,477	1,819,363	25,886	1.4
West:				
Alaska	550,043	561,276	11,233	2.0%
Arizona	3,665,228	3,754,666	89,438	2.4
California	29,760,021	30,597,578	837,557	2.7
Colorado	3,294,394	3,363,637	69,243	2.1
Idaho	1,006,749	1,029,283	22,534	2.2
Hawaii	1,108,229	1,129,170	20,941	1.9
Montana	799,065	818,348	19,283	2.4
Nevada	1,201,833	1,230,709	28,876	2.3
New Mexico	1,515,069	1,563,579	48,510	3.1
Oregon	2,842,321	2,896,472	54,151	1.9
Utah	1,722,850	1,753,188	30,338	1.7
Washington	4,866,692	4,958,320	91,628	1.8
Wyoming	453,588	463,629	10,041	2.2
U.S. Totals	248,709,873	252,730,369	4,020,496	1.6%

Source: U.S. Census Bureau, based on 1990 Post Enumeration Survey.

EFFECTS OF THE 1990 UNDERCOUNT

As noted previously, two of the more direct effects of census undercounts are that (1) they can affect the regional distribution of representatives from different states in Congress and (2) they can affect the interstate distribution of federal funds.

Effect of the Undercount on the House of Representatives

Article 1, Section 2, of the U.S. Constitution requires that the census be used to apportion seats in the U.S. House of Representatives (House). The effect of the 1990 undercount on the allocation of House seats to the 50 states is discussed below.

How Seats Are Allocated Among States. The U.S. Constitution provides that each state will have a minimum of one member in the House, and the current size of the House (435 seats) has not changed since the apportionment made following the 1910 census. Thus, the current apportionment calculation divides 385 seats (435 seats, minus the 50 seats automatically given) among the 50 states.

The method currently used for apportioning these 385 seats is called the “method of equal proportions.” It was adopted in 1941 following the 1940 census, and involves establishing a listing of the states according to “priority values.” These priority values are calculated using a formula which incorporates each state’s population growth relative to the size of each state. Seats 51 through 435 are assigned to the 50 states on the basis of this listing of priority values.

For example, following the 1990 census, each of the 50 states was given one seat out of the current total of 435. The next, or 51st seat, went to the state with the highest priority value (California) and thus became that state’s second seat. The state that had the next-highest priority value (New York) captured the 52nd seat, while the state with the third-highest priority value (California again) captured the 53rd seat. This process continued until all 435 seats had been assigned to a state.

California Would Have Gained an Additional Seat Absent the 1990 Undercount. Using the official 1990 census figures, California was allocated 52 seats in the House, with its 52nd seat being the 427th allocated under the priority ranking. Washington received the final 435th seat according to the priority ranking, and the next five states in priority order were Massachusetts, New Jersey, New York, Kentucky, and California (in that order). Thus, California qualified for the 440th seat, but because the total number of seats is fixed at 435, could not receive that 53rd seat.

Had the 1990 census undercount not occurred, the priority-order ranking would have been jumbled around. This is because the extent of the undercount differed by state. Adjusting for the undercount would have improved California’s priority ordering for its 53rd seat from number 440 to number 434, or inside the 435 House limit. Thus, California would have picked up a seat (its 53rd), due to its large relative undercount. As it turns out, this additional seat would have been at



the expense of Wisconsin, whose 9th seat would have slipped from priority-order ranking 429 using the official census data to 436 using the adjusted data. It should be noted that California is the only state in the nation to lose a seat because of the 1990 undercount. Moreover, the National Conference of State Legislators cited in a recent report an estimate by a demographic research firm that an adjusted 2000 census count would shift at least one seat to California as well.

Effect of the Undercount on Federal Funding Levels

Population helps determine the amount of federal funds states receive for a wide variety of public programs. Because of this, California's large relative census undercount caused it to receive

less than it should have under a wide range of federal formula grant programs throughout the 1990s. Precise dollar figures regarding federal funding effects are difficult to pinpoint, partly because some of the formulas which are used to distribute federal funds are very complex. Figure 3 lists California's 15 largest federal grant programs, and summarizes the eight that have been short-changed because of the undercount. Note that these figures are for a *single* fiscal year only, and are for only a fraction of the grant programs under which California receives federal funds. Extrapolating these figures for the entire decade suggests that the 1990 census undercount has likely cost California an estimated \$2.2 billion during the 1990s.

CALIFORNIA'S UNDERCOUNT— A GEOGRAPHIC PERSPECTIVE

Just like individual states experienced different relative 1990 census undercounts, so did California's different geographic regions. This reflects such factors as regionally different population characteristics, different living styles and arrangements, and different degrees of urbanization, suburbanization, and rural living.

Figure 4 (see page 12) shows the percentage undercounts which characterized the state's broad geographic regions in 1990. For this purpose, California's counties were grouped into five regions, each of which has somewhat different

and distinct economic and geographic characteristics. These regions include:

- ◆ **Southern California**, which consists of Los Angeles, Orange, San Diego, Ventura, Riverside, and San Bernardino Counties.
- ◆ The **San Francisco Bay Area**, which includes the nine counties which are proximate to the San Francisco Bay.
- ◆ The **Central Valley**, which ranges from Kern County in the south to Shasta

Figure 3

California's Estimated Loss in Federal Funding Due to 1990 Census Undercount Fifteen Largest Grant Programs

(In Thousands)

Federal Program	Amount ^a
Adoption Assistance	\$995 ^b
Prevention and Treatment of Substance Abuse	3,632 ^c
Child Care and Development	1,883 ^c
Employment and Training Assistance	— ^d
Employment Services	— ^d
Foster Care	9,353 ^b
Highway Planning and Construction	— ^d
Low Income Home Energy	— ^d
Medicaid	197,912 ^b
Rehabilitation Services	4,719
Social Services	3,213
Special Education	— ^d
Women Infants and Children Program (WIC)—Food	— ^d
WIC—Nutritional Services and Administration	— ^d
Vocational Education	1,128 ^c
Total	\$222,835

^a Federal fiscal year 1998 unless otherwise indicated.
^b Federal fiscal year 1997.
^c Federal fiscal year 1999.
^d These programs do not use population data to allocate funding.
 Source: United States General Accounting Office.

County in the north, and includes such midsized metropolitan areas as Sacramento, Stockton, Fresno, Modesto, and Bakersfield.

- ◆ The **Central Coast**, ranging from the counties of Santa Barbara in the south to Santa Cruz in the north.

- ◆ The **Rest-of-the-State**, which consists of the rural mountainous counties surrounding the Central Valley, and Imperial County.

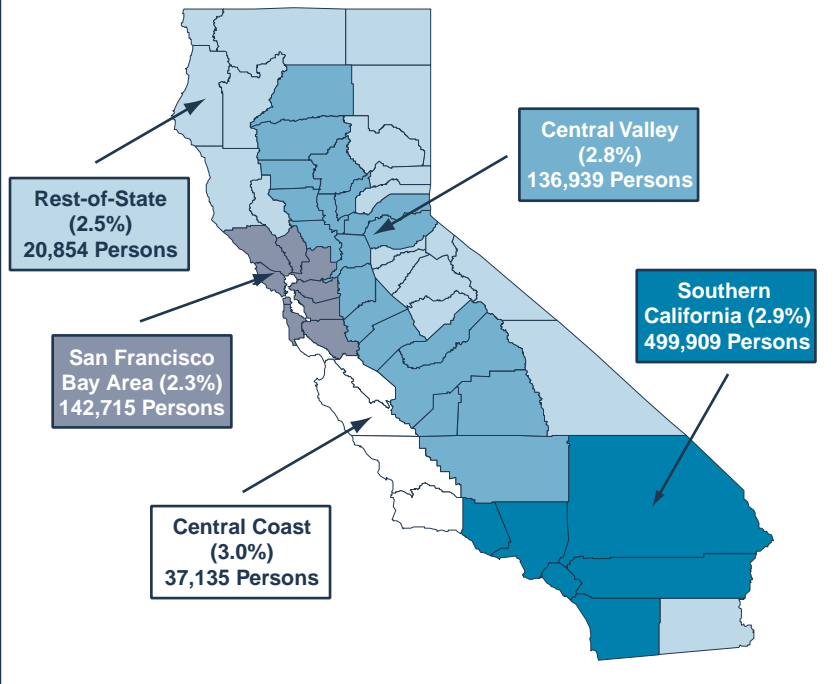
Figure 5 (see page 13) and Figure 6 (see page 15) report the undercount figures for California's counties and larger cities, respectively. Note that all but two counties (Marin and Placer) and three



Figure 4

California Census Undercount by Region

Persons and Percent
1990



cities (Santa Clarita, Thousand Oaks, and Torrance) had undercount rates higher than the nation as a whole (1.6 percent).

Regional Experience

The 1990 census undercount experience in the state's five broad geographic regions was as follows:

Southern California. In 1990, this region contained 57 percent of the state's population, and slightly more than 60 percent of the census undercount occurred in it. The experience for this region, however, is completely explained by Los Angeles County, which itself had an extremely

high undercount. In fact, Los Angeles was the only county in this region which experienced a higher undercount than the state in its entirety. It accounted for 30 percent of the state's population but was home to almost 37 percent (about 306,000) of the statewide's undercounted individuals.

San Francisco Bay Area. In contrast to Southern California, the nine-county San Francisco Bay Area accounted for less of the undercount than its population share—17 percent of the undercount, three percentage points below its share of 1990 official state population. Three counties—Marin, San Mateo, and Contra Costa—all posted

undercount figures considerably lower than the state (rates of 1.2 percent, 1.7 percent, and 1.8 percent, respectively). These three rates rank among the six lowest of all the counties, and offset higher undercount rates in such Bay Area counties as San Francisco and Alameda.

Central Valley. The Central Valley posted undercount figures which generally mirrored that of the state as a whole. That is, the region comprised 16 percent of the statewide population and accounted for 16 percent of the statewide undercount. However, like Southern California, the Central Valley also exhibited considerable

Figure 5

1990 California Census Undercount

Region/County	1990 Population Total		Estimated Undercount	
	Reported	Adjusted for Estimated Undercount	Amount	Percent
Southern California:				
Los Angeles	8,863,164	9,168,936	305,772	3.3%
Orange	2,410,556	2,461,397	50,841	2.1
Riverside	1,170,413	1,199,176	28,763	2.4
San Bernardino	1,418,380	1,455,650	37,270	2.6
San Diego	2,498,016	2,560,552	62,536	2.4
Ventura	669,016	683,743	14,727	2.2
Subtotals	17,029,545	17,529,454	499,909	2.9%
San Francisco Bay Area:				
Alameda	1,279,182	1,317,262	38,080	2.9%
Contra Costa	803,732	817,986	14,254	1.7
Marin	230,096	232,969	2,873	1.2
Napa	110,765	113,321	2,556	2.3
San Francisco	723,959	745,580	21,621	2.9
San Mateo	649,623	661,717	12,094	1.8
Santa Clara	1,497,577	1,531,401	33,824	2.2
Solano	340,421	348,548	8,127	2.3
Sonoma	388,222	397,508	9,286	2.3
Subtotals	6,023,577	6,166,292	142,715	2.3%
Central Valley:				
Butte	182,120	186,843	4,723	2.5%
Colusa	16,275	16,992	717	4.2
El Dorado	125,995	128,454	2,459	1.9
Fresno	667,490	692,182	24,692	3.6
Kern	543,477	558,924	15,447	2.8
Kings	101,469	105,195	3,726	3.5
Madera	88,090	91,267	3,177	3.5
Merced	178,403	185,469	7,066	3.8
Placer	172,796	175,290	2,494	1.4
Sacramento	1,041,219	1,065,246	24,027	2.3
San Joaquin	480,628	495,277	14,649	3.0
Shasta	147,036	150,146	3,110	2.1
Stanislaus	370,522	380,819	10,297	2.7

Continued



Region/County	1990 Population Total			
	Reported	Adjusted for Estimated Undercount	Estimated Undercount	
			Amount	Percent
Sutter	64,415	66,163	1,748	2.6
Tulare	311,921	323,772	11,851	3.7
Yolo	141,092	145,974	4,882	3.3
Yuba	58,228	60,102	1,874	3.1
Subtotals	4,691,176	4,828,115	136,939	2.8%
Central Coast:				
Monterey	355,660	367,820	12,160	3.3%
San Benito	36,697	38,192	1,495	3.9
San Luis Obispo	217,162	222,870	5,708	2.6
Santa Barbara	369,608	381,099	11,491	3.0
Santa Cruz	229,734	236,015	6,281	2.7
Subtotals	1,208,861	1,245,996	37,135	3.0%
Rest of State:				
Alpine	1,113	1,148	35	3.0%
Amador	30,039	30,482	443	1.5
Calaveras	31,998	32,606	608	1.9
Del Norte	23,460	24,035	575	2.4
Glenn	24,798	25,686	888	3.5
Humboldt	119,118	122,441	3,323	2.7
Imperial	109,303	113,271	3,968	3.5
Inyo	18,281	18,876	595	3.2
Lake	50,631	51,774	1,143	2.2
Lassen	27,598	28,162	564	2.0
Mariposa	14,302	14,673	371	2.5
Mendocino	80,345	82,788	2,443	3.0
Modoc	9,678	9,921	243	2.4
Mono	9,956	10,328	372	3.6
Nevada	78,510	79,826	1,316	1.6
Plumas	19,739	20,195	456	2.3
Sierra	3,318	3,401	83	2.4
Siskiyou	43,531	44,578	1,047	2.3
Tehama	49,625	50,823	1,198	2.4
Trinity	13,063	13,317	254	1.9
Tuolumne	48,456	49,390	934	1.9
Subtotals	806,862	827,721	20,859	2.5%
California Totals	29,760,021	30,597,578	837,557	2.7%

Figure 6

1990 California Census Undercount for Larger Cities^a

Region/City	1990 Census Total		Estimated Undercount	
	Reported	Adjusted for Estimated Undercount	Amount	Percent
Southern California:				
Anaheim	266,406	273,740	7,334	2.7%
Chula Vista	135,163	138,715	3,552	2.6
El Monte	106,209	110,792	4,583	4.1
Escondido	108,635	111,040	2,405	2.2
Fullerton	114,144	116,725	2,581	2.2
Garden Grove	143,050	146,412	3,362	2.3
Glendale	180,038	184,515	4,477	2.4
Huntington Beach	181,519	184,639	3,120	1.7
Inglewood	109,602	116,991	7,389	6.3
Irvine	110,330	112,191	1,861	1.7
Long Beach	429,433	445,925	16,492	3.7
Los Angeles	3,485,398	3,624,206	138,808	3.8
Moreno Valley	118,779	121,925	3,146	2.6
Oceanside	128,398	131,711	3,313	2.5
Ontario	133,179	137,458	4,279	3.1
Orange	110,658	112,738	2,080	1.8
Oxnard	142,216	147,164	4,948	3.4
Pasadena	131,591	136,431	4,840	3.5
Pomona	131,723	137,116	5,393	3.9
Rancho Cucamonga	101,409	103,309	1,900	1.8
Riverside	226,505	232,608	6,103	2.6
San Bernardino	164,164	170,249	6,085	3.6
San Diego	1,110,549	1,143,032	32,483	2.8
Santa Ana	293,742	305,815	12,073	3.9
Santa Clarita	110,642	111,997	1,355	1.2
Simi Valley	100,217	102,006	1,789	1.8
Thousand Oaks	104,352	105,407	1,055	1.0
Torrance	133,107	135,125	2,018	1.5

Continued



Region/City	1990 Census Total		Estimated Undercount	
	Reported	Adjusted for Estimated Undercount	Amount	Percent
San Francisco Bay Area:				
Berkeley	102,724	106,630	3,906	3.7%
Concord	111,348	113,137	1,789	1.6
Fremont	173,339	176,094	2,755	1.6
Hayward	111,498	114,720	3,222	2.8
Oakland	372,242	391,553	19,311	4.9
Salinas	108,777	112,703	3,926	3.5
San Francisco	723,959	745,573	21,614	2.9
San Jose	782,248	801,296	19,048	2.4
Santa Rosa	113,313	115,898	2,585	2.2
Sunnyvale	117,229	119,999	2,770	2.3
Vallejo	109,199	112,178	2,979	2.7
Central Valley:				
Bakersfield	174,820	179,398	4,578	2.6%
Fresno	354,202	366,527	12,325	3.4
Modesto	164,730	168,849	4,119	2.4
Sacramento	369,365	380,736	11,371	3.0
Stockton	210,943	218,358	7,415	3.4

^a Defined as cities with populations in excess of 100,000 as of 1990.

intercounty variation. The Central Valley is essentially comprised of two subregions—the San Joaquin Valley (extending from Kern County to San Joaquin County) and the Sacramento Valley (including Sacramento County up to Shasta County). The census data reveal that the *entire* San Joaquin Valley (all eight counties) suffered a collective undercount of 3.2 percent, significantly higher than the statewide average of 2.7 percent, and easily the highest of any subregion in the state. Conversely, the Sacramento Valley counties

posted the *lowest* undercount rate of any subregion—2.3 percent. This was despite the fact that it contains Colusa County, which had the highest undercount rate of all the counties in the state.

Central Coast and Rest-of-the-State. The final two broad geographic regions of California—the Central Coast and the Rest-of-the-State region—together posted undercount rates proportional to their shares of population. That is, the two regions collectively accounted for about 7 percent of the state’s official population count in 1990, as well as

about 7 percent of the estimated statewide census undercount. However, the Central Coast taken alone experienced the largest undercount rate of all five geographic regions (3 percent), driven by

large undercount rates in San Benito and Monterey Counties. These above-average rates were offset by generally lower undercount rates for the 21 remaining, mostly rural counties.

STATISTICAL SAMPLING— THE BUREAU'S PROPOSED SOLUTION

In response to the undercounting problem associated with past censuses—especially in 1990—the Census Bureau, as noted above, advocates the use of statistical sampling methods to increase accuracy. “Sampling” occurs whenever the information on a *portion* of the population is used to infer information on the population at *large*. This approach is intended to deal with both the component of the undercount problem associated with inaccuracies in the master list of households, as well as the component associated with imperfect information about the population residing in known households.

Actually, statistical sampling has been used since 1940 to obtain detailed demographic information about the population. In 1990, for example, about one-in-six residents were sent a special long-form questionnaire to fill out, the results of which were used to draw inferences about various attributes of the general population. (This same process with respect to the long-form questionnaire will be repeated in 2000.) Through 1990, however, the population *totals* themselves have reflected only the actual population *head*

count. The Census Bureau is proposing to change this traditional practice in 2000, and augment the head count itself by incorporating the results of sampling.

An Overview of the Proposal

The Original Proposal. The Census Bureau's original objective with respect to the upcoming 2000 census was to physically count the population in 90 percent of the households it was aware of in every census tract (each of which contains roughly 4,000 people). It would then account for the remainder of the population through scientific sampling techniques.

In addition to the above process, the Census Bureau wanted to conduct a second sample of 750,000 households nationwide drawn from all ethnic groups and geographic locations, as a sort of “quality assurance” check. This sample would allow the bureau's statisticians to gauge whether some particular demographic groups were miscounted in the first-round census calculations, in which case the preliminary results would then be adjusted accordingly.



According to the Census Bureau, if such a sampling method had been used, there was a 90 percent chance that its estimate of the nation's population would be within 0.1 percent of the true number. On the other hand, if no sampling methods were used, it estimated that the likely undercount would be about 1.9 percent, even higher than the 1.6 percent undercount in 1990.

The Current Plan. As discussed below, recent court decisions provide that congressional seats may not be apportioned using sample-adjusted data. This means that the Census Bureau cannot rely on a 90 percent coverage plan. Rather, it must try to make its coverage as close to 100 percent as possible. As a result, the bureau recently modified its original plan for sampling. It now plans to attempt to physically count everyone and then adjust this count using an Accuracy and Coverage Evaluation (ACE) survey involving approximately 300,000 households—twice as large as the one used in 1990 but less than half of the one originally proposed. The bureau believes that this modified plan, like the original plan, will significantly improve census accuracy—both in terms of identifying missed households and obtaining more accurate data on counted households. However, the unadjusted data will be used to apportion congressional seats, and the survey-adjusted data will be available for other purposes. The bureau estimates that the cost of conducting the 2000 census will be \$4.5 billion, \$1.7 billion higher than its original estimate, due to the need to try to count everyone.

A Detailed Look at How The 2000 Census Will Proceed

The initial phases of the Census 2000 project would be similar to those of previous census counts which did not use sampling. The first step calls for developing a list of every housing unit in the nation (the so-called master list). To contact all addresses, the Census Bureau plans to merge its 1990 Census Address List with a current address list from the U.S. Postal Service. Local governments would then be given the opportunity to review and update the list. The result will be a national listing consisting of about 120 million addresses. In April 2000, a series of mailings will be sent to each address on the list. Specifically, each address will be mailed a prenotice letter, followed by the official questionnaire, followed by a "reminder" or "thank you" postcard, as appropriate. To achieve as large a response rate as possible, census questionnaire forms also will be available in different languages at public places, such as libraries and post offices. Thus, the starting point for both the 1990 and 2000 censuses is in principle the same—a listing of the nation's households and other places of residences (such as nursing homes and dormitories).

As previously, the Census Bureau will focus on those households not responding to the census forms and/or other correspondence sent to them. Interviewers will go door-to-door to collect information from all the nonresponders they can locate, along with collecting information when possible about them from third-party sources such as neighbors and postal carriers.

There will still be persons who are missed or incorrectly enumerated in this process. For example, people in the responding housing units may make errors in filling out the census questionnaire. Likewise, some households may not even be included on the master-address list for the questionnaire. To address these problems, the bureau will undertake the post-census ACE survey. A similar survey was conducted in 1990, but the results were *not* incorporated into the final population figures. The ACE survey will be twice as big as the previous one and, hence, be more useful for adjusting the data.

In summary, the key difference between the 1990 census and the planned 2000 version is that in 2000 a population series will be available which incorporates the undercount identified by the post-census survey, if policymakers wish to use it.

What About the Homeless? In 2000, the Census Bureau will enumerate people at service locations (such as shelters, soup kitchens, and regularly scheduled food vans) that primarily serve people without housing. Efforts are also planned to enumerate persons without housing at targeted nonsheltered outdoor locations. In this way, the Census Bureau will seek to include people without housing in the census who might be missed in the traditional enumeration of housing units and group quarters.

Experience With Pre-Census "Trial Runs"

In preparation for the upcoming census, the bureau conducted "dress rehearsals" at three sites around the country in the spring of 1998. The

three dress rehearsal sites were Sacramento; the 11 rural counties surrounding Columbia, South Carolina; and the Menominee Reservation in Wisconsin. The three trial run sites were chosen because they have attributes reflective of the various challenges that Census 2000 will confront. Specifically, Sacramento was chosen to represent urban conditions; the multicounty area surrounding Columbia was selected because it provided an opportunity to test procedures in suburban and rural areas; and the final test site was chosen to demonstrate the special procedures planned for use on Native American Indian reservations.

The results from these three dress rehearsals will allow the bureau to evaluate the new procedures being considered for Census 2000. These include user-friendly forms and digital capture of forms. In addition, the Census Bureau tested the statistical sampling techniques it intends to use in the 2000 census. The site selection criteria for the dress rehearsal allowed populations to be assessed with certain attributes associated with the 1990 census undercount.

The Case of Sacramento. Sacramento was specifically selected because its population variations are felt to be reflective of those characterizing California generally, and it was felt that Sacramento provides a good "testing ground" to evaluate efforts to capture the classifications of persons generally missed in 1990. Figure 7 (see page 20) shows the results from the Sacramento trial run. These results suggest a continuation of two trends identified in the 1990 census.



- ◆ **The Undercount Problem Persists.** As the figure shows, the total undercount figure for Sacramento was 6.3 percent, more than double the reported undercount the city experienced in 1990.
- ◆ **Some Ethnic Groups Are Less Likely to Be Counted Than Others.** As the figure indicates, the undercount rates for each nonwhite subgroup exceeds that for the city as a whole. Indeed, African Americans, Native Americans, and the “Other”

category all experienced undercount rates exceeding 8 percent, and the composite rate for non-Whites was 7.7 percent. In contrast, the undercount rate for Whites was 4.9 percent.

Thus, many experts believe that the results of the Sacramento rehearsal suggest that, absent statistical sampling as a corrective remedy, the census undercount resulting from the 2000 census may be even larger and the population totals more problematic than for the 1990 experience.

Figure 7

Summary Results of 1998 Census Dress Rehearsal (Sacramento)

Ethnic Group	Population Totals			
	Unadjusted for Undercount	Adjusted for Undercount	Undercount	
			Amount	Percent
White	185,478	195,046	9,568	4.9%
Black/African American	58,443	63,826	5,383	8.4
Asian/Pacific Islander	59,265	63,125	3,860	6.1
American Indian, Alaskan Native	11,270	12,327	1,057	8.6
Other	63,285	68,988	5,703	8.3
Totals	377,741	403,312	25,571	6.3%

WHERE DOES THE SAMPLING CONTROVERSY STAND TODAY?

Litigation Challenging Sampling

In 1998, in response to suits challenging the use of sampling for census purposes, it was ruled at the federal district court level that sampling methods may *not* be used to produce the population counts *used to reapportion seats* in Congress.

These rulings were upheld by the Supreme Court in January 1999, when it found that the census law directly prohibits use of statistical sampling to adjust population figures used to allocate House of Representative members among the states. The court drew a distinction, however, between using sampling to adjust the head-count figures used to apportion seats in the House among the 50 states, and statistical adjustment of those figures for other purposes (such as the distribution of federal funds to the states). While federal law bars sampling for apportionment, the court said it permits and perhaps even may require statistical adjustments for other purposes.

Thus, the court's interpretation of the Census Act suggests that population counts adjusted by sampling could or even should be used for these other purposes. Both Clinton Administration officials and Census Bureau officials have signaled their intention to do so.

Given the court decision, the Census Bureau plans to produce two sets of population figures—a traditional head-count version for the purpose of congressional apportionment, and then a second set of numbers which corrects for the undercount. The latter, more complete figures would be made available in a form that allows them to be used, if so desired by policymakers, for intrastate redistricting, determining the allocation of federal funds, and various other purposes. However, this would be contingent on Congress agreeing to appropriate the money for the Census Bureau to produce sample-adjusted figures following the regular head-count enumeration.



WHAT COMES NEXT?

At this point, the 2000 census and the sampling controversy surrounding it remains an unfinished story. Several key issues remain to be resolved.

Federal Issues. At the federal level, there are two key decision points. The first involves whether Congress will fund the ACE survey. If it does *not*, that will be the end of the story, as only one set of population figures will be produced—reflecting an actual census headcount that is unadjusted for the undercount through sampling. If the ACE survey *is* funded, however, a second key decision will then have to be made—namely, what set of population data should be used to distribute federal funds amongst the states, the unadjusted or sample-adjusted census results?

California Issues. Should sample-adjusted census data be made available through the ACE survey, the state will have to face several important issues. The first involves redistricting—specifically, which set of population data (adjusted

versus unadjusted) should be used to re-draw the boundaries of the state’s Congressional districts, as well as the Legislature’s Senate and Assembly districts?

The second key California issue facing the Legislature will involve the geographic dispersion of certain state funds to localities. Under current law, for example, population influences how vehicle license fee revenues, certain gasoline tax proceeds, and funds under the Citizens’ Option for Public Safety (COPS) program are geographically allocated. Thus, the amounts of dollars going to different localities under these programs will depend, in part, on whether adjusted versus unadjusted population figures are used.

Thus, depending on actions at the federal level, the state may soon have to deal with the impact of sampling on the census data.



Acknowledgments

This report was prepared by Robert Ingenito, under the supervision of David Vasché. The Legislative Analyst's Office (LAO) is a nonpartisan office which provides fiscal and policy information and advice to the Legislature.



LAO Publications

To request publications call (916) 445-2375. This report and others, as well as an E-mail subscription service, are available on the LAO's internet site at www.lao.ca.gov. The LAO is located at 925 L Street, Suite 1000, Sacramento, CA 95814.