A KIDS COUNT/PRB Report on

CENSUS 2000

The Undercount in the 2000 Census

By Barry Edmonston

The Annie E. Casey Foundation and
The Population Reference Bureau
May 2002

PRB

KIDS COUNT

KIDS COUNT, a project of the Annie E. Casey Foundation, is a national and state-by-state effort to track the status of children in the United States. By providing policymakers and citizens with benchmarks of child well-being, KIDS COUNT seeks to enrich local, state, and national discussions concerning ways to secure better futures for all children. At the national level, the principal activity of the initiative is the publication of the annual KIDS COUNT Data Book, which uses the best available data to measure the educational, social, economic, and physical well-being of children. The Foundation also funds a nationwide network of state-level KIDS COUNT projects that provide a more detailed community-by-community picture of the condition of children.

The Population Reference Bureau (PRB)

Founded in 1929, the Population Reference Bureau is the leader in providing timely, objective information on U.S. and international population trends and their implications. PRB informs policymakers, educators, the media, and concerned citizens working in the public interest around the world through a broad range of activities, including publications, information services, seminars and workshops, and technical support. PRB is a nonprofit, nonadvocacy organization. Our efforts are supported by government contracts, foundation grants, individual and corporate contributions, and the sale of publications.

KIDS COUNT/PRB Reports on Census 2000

This paper is part of a series of reports on the 2000 Census prepared for the nationwide network of KIDS COUNT projects. These reports have been guided by the recommendations of an expert advisory group of data users and child advocates, brought together in a series of meetings by the Annie E. Casey Foundation and the Population Reference Bureau. Members of the advisory group have provided valuable assistance about how to interpret and use data from the 2000 Census.

A list of the advisory group members can be found at the back of this report.

For more information or for a pdf version of this report, visit the Annie E. Casey Foundation's KIDS COUNT Web site at www.kidscount.org or PRB's Ameristat Web site at www.ameristat.org.

The Undercount in the 2000 Census

By Barry Edmonston
Portland State University

The Annie E. Casey Foundation and
The Population Reference Bureau
May 2002

EXECUTIVE SUMMARY

Every 10 years, the national census provides a snapshot of the number of U.S. residents and their characteristics. Census data are important for determining political representation, allocating federal and state funds, and monitoring social and economic trends. It is well-known, however, that the decennial census is subject to error. Decades of research have shown that the decennial census, like population censuses in other countries, fails to count all residents, and that census undercount rates differ by age, sex, and race. The 2000 Census was no exception. Despite the best efforts and careful planning of Census Bureau staff, the direct, physical enumeration of the U.S. population fell short of independent estimates of the true U.S. population. This report summarizes what is known about census undercount and the methods used to adjust census data, with an emphasis on the undercount of children and minorities.

- The Census Bureau reported that more than 2.0 million children were missed in the 1990 Census, accounting for more than half of the net undercount.[1] In 2000, there was a net undercount of about 340,000 people. This consisted of a net *overcount* of about 150,000 people ages 18 and older, and a net *undercount* of about 490,000 children.

- In both the 1990 and 2000 Censuses, the net undercount rate was higher for people who were American Indian, Asian, black, or Hispanic than it was for the overall population

- The undercount can cause inequities in the distribution of federal and state funds, and can result in poorly targeted programs for the provision of social services for children and families.

- Children in mobile and complex households, children living with grandparents or other relatives, foster children, and children with a divorced or never-married parent are more likely to be missed in the census.

- In March 2001, a decision was made not to adjust census results for the purpose of redistricting congressional and other political districts.[2] In October 2001, the Census Bureau announced that it also would not adjust 2000 Census data for nonpolitical purposes, such as allocation of federal funds.[3]

- Whatever the other advantages or disadvantages of adjustment for census undercount, the decision not to adjust the 2000 Census simplifies comparisons between 1990 and 2000 Census data.

TRENDS AND REASONS FOR CENSUS UNDERCOUNT

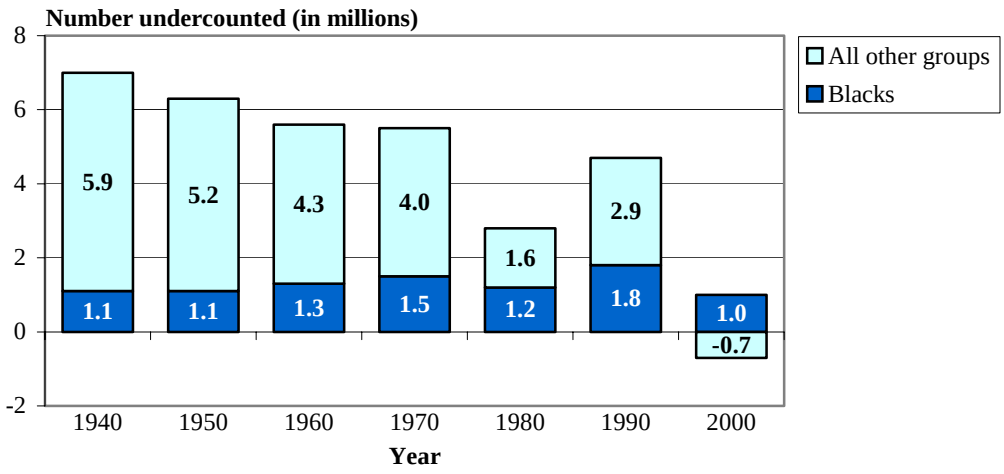
Although the census count of the U.S. population has never been complete, public concerns about its incompleteness have increased in recent decades. The census is the sole basis for apportionment of congressional seats and is relied on heavily for the distribution of federal funds. Improved statistical and demographic techniques permit the Census Bureau to estimate the incompleteness of the census with more accuracy than in the past. Thus, concern about census incompleteness springs, ironically, from the improved professional work of Census Bureau staff and the public's extraordinary expectations for a "complete" census count.

Coverage estimates, which measure the extent to which the census counts all the people, are made using two methods. One method is to conduct a larger sample survey in conjunction with the decennial census, match all individuals in the survey to those reported in the census, and then estimate the number of unenumerated people in the census by age, sex, and race. The 2000 Accuracy and Coverage Evaluation (ACE) is an example of such a survey. The second method, demographic analysis, uses birth and death records for previous years, immigration and emigration data, and previous censuses to develop an estimate of the population independent of the census.

Demographic analysis of coverage shows that the net national undercount (the number of people omitted minus the number overcounted) was about 7.0 million in 1940, 6.3 million in 1950, 5.6 million in 1960, 5.5 million in 1970, 2.8 million in 1980, 4.7 million in 1990, and 340,000 in 2000 (see Figure 1). There are different trends for the black population and the rest of the population (referred to as "nonblacks" in Census Bureau studies). Because of the lack of consistent birth, death, and immigration data by race and ethnicity, demographic analysis cannot provide undercount estimates for American Indians, Hispanics, Asians, and Pacific Islanders;

demographic analysis provides undercount estimates only for the black, nonblack, and total populations. The undercount for blacks exceeded 1.0 million in the 1940 and 1950 censuses, rising to 1.8 million in 1990, and decreasing to about 1.0 million in 2000. Net undercount for all other population race groups combined (that is, nonblacks) steadily declined from 5.9 million in 1940 to 1.6 million in 1980, increased to 2.9 million in 1990, and dropped to a net *overcount* of 700,000 in 2000.

Figure 1: Net Undercount of Blacks and All Other Population Groups, 1940-2000

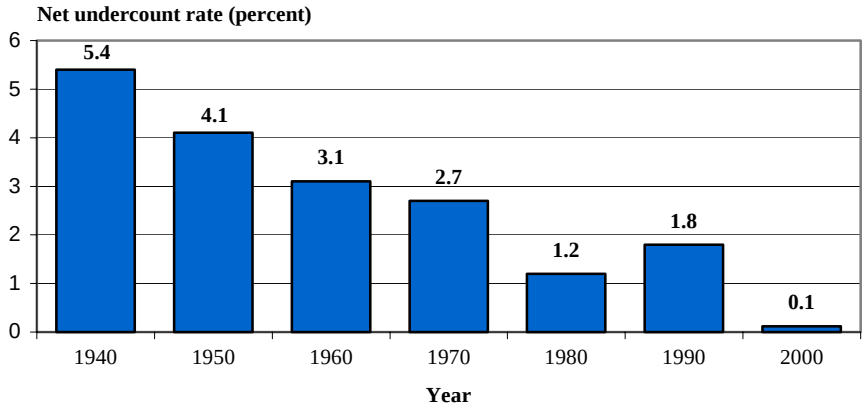


Sources: Edmonston and Schultze, eds., *Modernizing the U.S. Census* (1995): table 2.1; and Robinson "ESCAP II: Demographic Analysis Results" (2001): table A.

As Figure 2 shows, the undercount rate dropped steadily from 1940 (5.4 percent) to 1980 (1.2 percent), before rising in 1990 (1.8 percent) for the first time in 50 years, and then declining to 0.1 percent in 2000. Because the net undercount rate has been decreasing for both blacks and nonblacks over time, the difference between the black and nonblack rates has decreased only modestly since 1940 (see Figure 3). In 1940, the undercount rate for blacks was 3.4 percentage points higher than the nonblack rate. The difference between the two rates fluctuated before

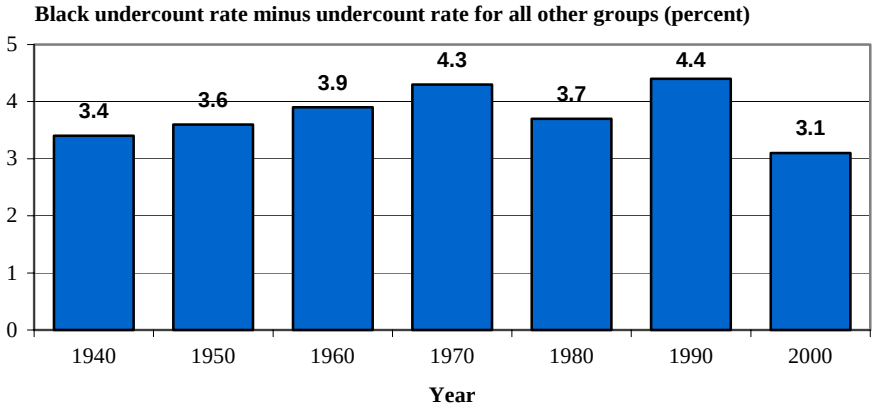
reaching a high of 4.4 percentage points in 1990. The difference between the black and nonblack rates declined to 3.1 percentage points in 2000, suggesting that, even with an expensive census that produced a net overcount of the nonblack population, it is very difficult to reduce the differential undercount by race using the traditional census design.

Figure 2: Net Population Undercount Rates, 1940-2000



Sources: Edmonston and Schultze, eds., *Modernizing the U.S. Census* (1995): table 2.1; and Robinson "ESCAP II: Demographic Analysis Results" (2001): table A.

Figure 3: Difference in the Net Undercount Rate Between Blacks and All Other Population Groups, 1940-2000



Sources: Edmonston and Schultze, eds., *Modernizing the U.S. Census* (1995): table 2.1; and Robinson "ESCAP II: Demographic Analysis Results" (2001): table A.

Undercount of Children and Minorities in the 1990 Census

Census Bureau estimates of 1990 undercount rates for the population under 20 years of age were about twice the rate for the adult population. Net census undercount rates are highest for children under 5 years of age, and diminish as young people become older, although undercount rates for teenagers are higher than the national average for all persons. The undercount rates in the 1990 Census were 3.7 percent for children ages 0 to 4, 3.4 percent for children ages 5 to 9, 3.1 percent for children ages 10 to 14, and 2.9 percent for children ages 15 to19 (see Table 1), compared with a net census undercount rate of 1.6 percent for the overall U.S. population.[4]

Table 1. 1990 Census Undercount Rates by Sex and Race Group, for the Population Under 20 Years of Age, United States

			Race Group				
	Age	Total	American Indian	Asian	Black	Hispanic	White[1]
Both sexes							
	0–4	3.7	7.6	4.3	8.3	6.0	2.2
	5–9	3.4	7.0	3.6	7.9	5.4	2.1
	10–14	3.1	6.7	3.1	7.6	5.0	1.9
	15–19	2.9	6.5	4.4	6.0	5.7	1.7
Males							
	0–4	3.7	7.5	4.3	8.3	6.0	2.2
	5–9	3.3	7.0	3.6	7.8	5.4	2.0
	10–14	3.1	6.7	3.1	7.5	4.9	1.9
	15–19	2.9	6.6	6.4	5.5	6.0	1.6
Females							
	0–4	3.7	7.7	4.3	8.3	6.0	2.2
	5–9	3.4	7.0	3.6	7.9	5.4	2.1
	10–14	3.2	6.6	3.1	7.6	5.0	1.9
	15–19	2.9	6.5	2.3	6.6	5.3	1.7

[1]The white population includes children who were identified as white and non-Hispanic.

Source: U.S. Census Bureau, “Net Population Adjustment Matrix Data File” (1996).

Net census undercount rates in 1990 were highest for black and American Indian children, with rates above 6 percent for all age groups. Undercount rates were also higher for Asian and Hispanic younger people than for the white population. The undercount of Asian and Hispanic groups is likely to have been influenced by the relatively large numbers who are foreign-born—many of whom may not have understood census questionnaires and procedures.

Net undercount rates were similar for males and females, with a few exceptions: Asian and Hispanic males ages 15 to 19 had considerably higher undercount rates than comparable females, and black females ages 15 to 19 had higher rates than black males in the same age group.

About 40 percent of the undercounted children and young people in 1990 were identified as white (see Table 2). Most of the remaining undercount for the population under 20 years of age were black (about 35 percent) or Hispanic (20 percent). Because of their relatively small numbers in the U.S. population, American Indian and Asian children accounted for a smaller proportion of the undercounted population.

Table 2. Distribution of the Undercounted Population Under 20 Years of Age in the 1990 U.S. Decennial Census, by Race and Age Groups

		Race Group				
Age	Total	American Indian	Asian	Black	Hispanic	White[1]
0-4	100.0	2.3	3.7	34.4	21.3	38.4
5-9	100.0	2.3	3.6	34.6	19.5	40.1
10-14	100.0	2.3	3.1	36.5	18.4	39.6
15-19	100.0	2.3	5.3	31.6	22.9	37.9

[1]The white population includes children who were identified as white and non-Hispanic.

Source: U.S. Census Bureau, “Net Population Adjustment Matrix Data File” (1996).

Evidence From the 2000 Census

Undercount results from the 2000 Census became available in October 2001 in a report that presented revised demographic analysis estimates as well as results from the Accuracy and Coverage Evaluation (ACE) survey.[5] Although undercount estimates from demographic analysis and surveys have been generally consistent for previous censuses, the two sets of estimates for the 2000 Census differ in several important ways. First, there are differences in the overall levels of undercount: Demographic analysis reported a net undercount of 0.12 percent, while the ACE survey reported a net undercount of 1.15 percent. Second, there are conflicting estimates of the black/nonblack net undercount differences. Demographic analysis suggested that the net undercount for the black population was 3.07 percentage points higher than for the nonblack population. Survey results from the ACE, however, reported that the undercount rate for blacks was only 1.06 percentage points higher than for the nonblack population. In other words, demographic analysis suggested that there was very little overall net undercount in the 2000 Census, but that a substantial racial difference in undercount persisted. The ACE survey, on the other hand, implied that there was a modest overall net undercount in the 2000 Census, but that racial differences in undercount were dramatically reduced, compared with previous censuses.

For the population under age 18, both demographic analysis and the ACE survey found that there was some undercount, although it was not as large as in the 1990 Census. Demographic analysis indicates undercount rates of 0.45 percent for male children and 0.89 percent for female children. The ACE survey found that the net undercount for the population under age 18 was virtually the same for males and females, with rates of 1.53 percent for males and 1.54 percent for females.

Both demographic analysis and the ACE survey found racial differences in the net undercount of children. Demographic analysis uncovered a racial difference of 0.76 percentage points, with net undercount rates of 1.30 percent for black children and 0.54 percent for nonblack children. The ACE survey reported a larger racial difference of 1.65 percentage points, with net undercount rates of 2.92 percent for black children and 1.26 percent for nonblack children.

Net undercount rates for the 2000 Census, according to the ACE survey, continue to be higher for racial minorities than for whites.[6] The net undercount rate for the Asian population was 0.96 percent, slightly lower than the overall national average. The net undercount rates for other groups, however, were higher than the national average: 2.85 percent for Hispanics, 3.28 for American Indians off reservation and 4.74 percent for those on reservation, and 4.60 for Native Hawaiian and other Pacific Islanders.

Reasons for Census Undercount of Children

Some children were missed in the 1990 Census because they lived in housing units that were never identified or in families that did not submit a completed form; about half of the overall 1990 Census undercount is attributable to missed housing units. Other children were missed because of undercoverage within housing units.

Undercount of Housing Units. A special goal for the 2000 Census was to reduce the number of missed housing units. Past censuses have been criticized by local governments that believed that some housing units had not been counted by the census. In the 1990 and previous censuses, however, the Census Bureau had deemed the address list confidential, so local governments were not able to make direct comparisons between their address list and the one used for the census. In the 1990s, Congress passed legislation that allowed the Census Bureau to share addresses with

local governments. This improved the 2000 Census address list and reduced local governments' criticism of the census enumeration.

The Census Bureau also worked closely with the U.S. Postal Service to improve the residential address list used in the 2000 Census. As in previous censuses, the Postal Service provided updated addresses that were being used by letter carriers. In addition, the Postal Service, in cooperation with local governments, continued to convert rural addresses (that is, rural routes and box numbers, post office boxes, and general delivery) to city-style addresses (that is, street names and numbered addresses). Such conversion allowed the Postal Service to more efficiently sort and deliver mail and the Census Bureau to handle the automated matching for assignment of geographic codes. Finally, the Postal Service assisted the Census Bureau in identifying vacant housing units for the 2000 Census.

The goal of these partnerships with local governments and the Postal Service was to reduce the undercount of housing units in the 2000 Census. To be sure, some housing units were still missed in the 2000 Census, but the number of missed units was likely to be considerably smaller than in previous censuses.

Undercoverage of Children Within Housing Units. Even when the census identifies a housing unit (either when someone in the housing unit completes and mails back a questionnaire or when a census enumerator canvasses the household), data on people in the housing units are not always correctly reported.

Undercoverage of children is more common in mobile and complex households. Mobile households include people who are temporary residents or who frequently change their residence. In mobile households that include college students, for example, it may be difficult to identify how many people are actually residing in the housing unit at the time of the census.

Foster children, children living with grandparents or other relatives, and children whose parents are divorced are also more likely to be missed, because it is harder to identify a usual place of residence.

Examples of complex households include unrelated individuals (such as boarders or roommates), people who have an ambiguous family status (for example, the cousin of an ex-spouse), or households with two or more separate families (such as several families living in a single housing unit in order to share the rent). Complex households contribute to census undercoverage because it is difficult to correctly identify the erratic or irregular membership of the household members.

Other Reasons for Undercount of People in Households. The 1990 Census sponsored 29 ethnographic studies throughout the country in order to learn more about the factors contributing to census undercount.[7] The studies focused on ethnic and racial minorities and point to a set of factors that contribute to differential census undercount:

- **Fear of government and outsiders.** When residents fear outsiders, they are more likely to conceal information from the census enumerators. The sources of fear vary, ranging from fear of arrest for criminal activity to a general distrust of government. For those who distrust the confidentiality of the census, education and outreach could be used to try to reassure the public that census data are strictly confidential. For example, the Census Bureau made special efforts in the 2000 Census to encourage participation of undocumented immigrants. There is little the Bureau can do, however, to improve the collection of data for households involved in criminal activity.
- **Limited knowledge of English.** Language barriers were identified as an important source of undercoverage in the 1990 Census. People who were illiterate or who did not

know English were less likely to complete the census questionnaires. The 2000 Census included major improvements to increase the participation of non-English speakers in the census, including census questionnaires in six languages, census guides in almost 50 languages, and partnerships with local ethnic associations.

- **Mobile people and households.** It is difficult for the census to count people who move frequently, since they may be moving at the time of the census or it may be difficult to establish where they are living on census day. New immigrants, who may move frequently in search of employment or cheaper rental housing, and college students can be particularly hard to find.
- **Irregular household members.** People who have an irregular relationship to other household members are less likely to be counted. The census questionnaire asks for one person to be identified as "person 1"; that person is then referred to as the head of household. People who are related to "person 1" are fairly easy to identify because they have some regular relationship, such as spouse, daughter, or father. People who have no formal relationship to "person 1," however, are more difficult to capture on the census form and thus are more often missed in the reporting.

Undercoverage of Special Populations. The factors listed above may interact with each other, increasing the undercoverage rates for special populations. Recent immigrants, for example, are more likely to be mobile, have complex household arrangements, and lack English-language skills. For such households, it becomes challenging to identify the households, obtain full information on household members, and include any irregular household members. Undercoverage is a problem for certain rural populations as well. Poor families in rural areas often lack city-style addresses and may be distrustful of attempts to count the members of their

families. Illiteracy is another factor that leads to undercoverage. Seasonal and migrant farmworkers are special population groups that have been undercounted to a great extent in previous censuses. Farmworkers tend to be mobile, are sometimes undocumented residents of the United States, may reside in irregular household arrangements, and often do not speak English—all factors that are associated with higher levels of undercount.

WHY UNDERCOUNT MATTERS

Fiscal Implications of Undercount

The undercount affects the distribution of certain federal and state funds that are allocated on the basis of population. Funds for education, health, transportation, housing, community services, and job training are allocated to geographic areas on the basis of population size and social and economic factors. In fiscal year 1998, the federal government disbursed about $185 billion to state and local governments, using formulas involving census data. Of the federal programs that distribute funds based on population counts, Medicaid is the largest, followed by the Highway Planning and Construction Program, Title 1 Grants to Local Education Agencies, Foster Care, and Federal Mass Transit Grants. Several studies have examined the effect of adjusting for census undercount on the distribution of funds to state and local governments. All studies of the 1970, 1980, and 1990 Censuses, for example, concluded that the impact of such an adjustment would have been relatively small.

The federal allocation of grants involving census counts in 1998—for 22 large formula grant programs—was $162 billion, or about $599 per capita for eligible population jurisdictions.[8] However, there are several reasons why adjusting the allocation for undercount would not simply result in an additional $599 per net undercounted person. First, population is only one of several factors used in many federal formula grant allocations; in such programs, an increase of

population results in only a partial increase in funding. Second, although many grant allocations increase with population gains, there are some programs (such as the Community Development Block Grant Program in the U.S. Department of Housing and Urban Development) in which funding is *reduced* with population growth. Third, and most important, federal grant formulas are largely fixed in their total amount. In reality, as estimates of the total U.S. population are increased by adjusting for estimated undercount, smaller amounts of funds per capita would be available for allocation. If, for example, a fixed sum were apportioned among geographic areas on the basis of population size alone and the population of every geographic area doubled, there would be no change in funds allocated to any area, only a reduction in the per capita amount.

If 1998 obligations for each state and local jurisdiction count were corrected for estimated undercount, some states would have lost while others would have gained. The overall amount per net undercounted person among gaining areas would have been about $224, considerably less than the average per capita obligation ($599). Although 27 states and the District of Columbia would have gained because of adjusted population counts, federal funding for the remaining 23 states would have been reduced. Among the gaining states, federal obligations for both California and Texas would have increased by more than 1 percent; among losing states, federal obligations would have decreased by more than 1 percent for Michigan, Minnesota, Pennsylvania, and Rhode Island. Many governments with only modest population undercounts would not, in fact, have actually gained additional federal grant monies. As a result, the effect on redistribution of federal funds would have been modest: Only 0.3 percent, or about $490 million, of the total 1998 federal obligations would have been altered by adjusting the population count.

By definition, the distribution of money under these programs would change if there were a differential change in the population count. The effect of the undercount on each state’s share of

a fixed total of distributed funds depends on state characteristics. Moreover, the amount of money gained and lost is obviously related not only to the estimated undercount rate, but also to population size.

However, in federal funding allocation programs, social and economic factors as well as population counts are used to determine how funds are distributed. The use of these other factors points to the importance of enumeration and of accurate data for optimal program planning and equitable distribution of funds. Reducing the undercount and improving the accuracy of collected data are both important to the Census Bureau, which needs to provide accurate data to ensure the fair funding of federal programs.

Political Implications of Undercount

Underenumeration in the census has serious political, economic, and social implications. The decennial population count reported in the census affects the state apportionment of seats in the U.S. House of Representatives, as well as the geographic boundaries for congressional districts, state legislative districts, and local political limits, such as city council districts. Under the "equal proportions" methods for federal apportionment, a shift of even a relatively small number of people could result in a change in a state's representation.

An estimate of what the 2000 Census results would have been, had statistical methods been used to adjust the count, can be obtained by using Census Bureau survey information to make a "correction" for the undercount of each state's population. It is difficult to determine precisely how the application of corrections for the estimated undercoverage in the 2000 Census would have affected actual congressional reapportionment, because the adjustment would have been done for small geographic areas. But if "coverage correction factors" released by the U.S.

Census Bureau in March 2001 had been applied to each state's population data, Texas probably would have gained a congressional seat and Ohio would have likely lost a seat.[9]

Congressional redistricting would be affected to a greater extent than apportionment because virtually all congressional districts, except for those in states with single districts, would have their boundaries changed by adjusted census block data. Moreover, a census that is adjusted for undercoverage in the physical enumeration would affect the redistricting of state legislatures and city councils that rely on decennial census data.

There are also several implications of the undercount for minority groups. In political representation and funding that is based on population, undercounted groups get less credit for their population than they are due. Political districts for undercounted areas, drawn relative to population, are "overpopulated" (that is, they have more people than the official data indicate), compared with accurately counted districts. "Overpopulated" districts result in underrepresentation of minority areas (that is, fewer districts than should be the case) at all levels of government—federal, state, and local—in which political representation is determined based on population size. For all state and local districts, the possibility that undercoverage will affect a district's boundaries depends on the size of the district; the coverage rates by age, sex, and race; the distribution of the population by age, sex, and race; and the undercoverage rates of contiguous districts.

EFFORTS TO REDUCE THE UNDERCOUNT

Past Efforts to Reduce Undercount Through Intensive Enumeration

The traditional census approach begins with construction of an address register, including elaborate procedures to improve coverage of the population. Census forms are then mailed to a

comprehensive list of residential addresses, with instructions to mail back the completed questionnaire. Not all households return their completed questionnaires within a reasonable period of time. For households that do not respond to the main questions about the number of family members and their key demographic characteristics (33 percent of housing units in 2000), census enumerators undertake an intensive follow-up effort to determine whether the unit is occupied and, if so, to contact the household and obtain responses. Repeat visits are made, administrative records are sometimes examined, and special programs to contact particular groups (for example, the homeless; people in institutions, dormitories, or barracks; other people who do not live in regular household settings) are carried out. The process is continued for an extended period of time in an effort to physically enumerate each household and its occupants.

The Census Bureau has made special efforts to improve coverage in recent censuses, including, for example, a follow-up of people reporting a change of address to the U.S. Postal Service during the census enumeration period; a campaign to find people who were missed in the census by contacting community organizations or visiting places frequented by transients; and an effort to match administrative records to census lists for selected areas. These efforts are frequently expensive, both in absolute terms and in terms of the cost per person or housing unit. The results from the returned mail questionnaires, enumerator follow-up, and coverage improvement efforts are combined to produce the actual census count of the U.S. population, reported by the Census Bureau for the nation and for subdivisions down to the block level.

The traditional approach, direct enumeration, has had two basic problems: high and rapidly rising costs, and high differential undercount. In fact, many analysts believe that the traditional approach has been pushed well beyond the point at which it adds to the overall accuracy of the

census count. The costs of taking the census have risen sharply, even after allowing for inflation and population growth.[10]

Furthermore, the 1990 Census produced a net undercount of 1.6 percent for the nation as a whole, and 3.2 percent for children; the 2000 Census produced a net undercount of 0.1 percent for the entire population, and 0.7 percent for children. These undercount rates included overcounting in some areas and among some groups—which was more than offset by undercounting among other areas and groups. Blacks and Hispanics, Asians and Pacific Islanders, American Indians and Native Alaskans, renters, and residents of poor inner-city areas were undercounted by larger percentages than the nation as a whole. Currently, there are no new procedures available within the traditional census approach that could substantially reduce differential undercount. Historically undercounted population groups experienced some coverage improvements through a more intensive, expensive traditional census in the year 2000, but substantial differential undercount by age, sex, and race continued to exist.

Efforts to Reduce Undercount Through Statistical Adjustment

As early as the 1940 Census, demographers compared Selective Service registration records with state-level census figures for males by age and race, and concluded that there was a national undercount of young adult men in the 1940s.[11] Their work was a direct antecedent for the continuing tradition of administrative records research at Statistics Canada, where reverse record linkage has been used since 1976 to estimate the undercount in Canadian censuses. In a pioneering study published in 1955, Ansley Coale showed how demographic analysis could be used to reveal undercount rates by age, sex, and race.[12] By the end of the 1960s, the U.S. Census Bureau was making estimates of undercoverage in the 1960 census.[13] The work of Census

Bureau statisticians and demographers in following years provided evidence that the census counts for groups such as African Americans were quite defective.

By the 1970s, demographers had developed a set of methods for reconstructing the U.S. population by age, sex, and race. A panel convened by the National Academy of Sciences reviewed the issue of census undercount and, for the first time, advocated a set of approaches that could be used to adjust census counts.[14] Although this panel of experts stopped short of endorsing an adjusted census count, they did advocate steps to examine the feasibility of adjusting census counts, with suggestions for evaluating the 1980 census.

By the mid-1980s, there was considerable discussion about the issue of statistical adjustment of decennial census counts. Census Bureau staff had completed several papers examining the issues of statistical adjustment, and the methods received widespread attention. At the same time, National Academy of Sciences panels endorsed efforts to adjust the 1990 Census counts. Nevertheless, political controversies increased, generating strong arguments for and against adjustment of census data.

For the 1990 Census, the U.S. Census Bureau conducted a large, independent household survey called the Post-Enumeration Survey (PES) in order to measure differential undercount by age, sex, race, and several other individual and housing characteristics. Analysis of PES results after the 1990 Census indicated that 1.6 percent of the population was missed in the direct enumeration. The 1990 net undercount consisted of about 8.4 million people who were missed, combined with about 4.4 million people who were “double-counted,” yielding a total net undercount of about 4.0 million people. Dividing this 4.0 million by the estimated “true” population (253 million) yields the net undercount rate of 1.6 percent.

Because of concerns about the decision to adjust the 1990 Census counts, an advisory group was formed to provide independent advice to the secretary of commerce. The formation of the advisory group itself was controversial, and partisan nominations split the group into advocates for and against adjustment. U.S. Census Bureau staff recommended that the PES be used to adjust the final count, but the secretary of commerce's advisory group split their vote along partisan lines. In the end, the secretary of commerce decided not to adjust the final count, and the numbers collected by direct enumeration stood as the final census count.

For the 2000 Census, the Census Bureau developed the Accuracy and Coverage Evaluation (ACE) survey. Although it differed in many details from the PES used in the 1990 Census, it had the same purpose: to evaluate census undercount and, potentially, to be used to adjust for undercount. The Census Bureau and outside advisory groups, including several panels of experts convened by the National Academy of Sciences, concluded that the ACE was statistically sound and operationally feasible (See Appendix A for a description of how the ACE worked in the 2000 Census).

Following the 2000 Census, two problems with the measurement of undercount emerged. First, demographic analysis and the ACE survey produced different estimates, in both absolute numbers and relative rates, of the undercount. This divergence created uncertainty about the reliability of the undercount estimates and raised questions about the merits of using the ACE undercount estimates—by age, sex, race, and other demographic characteristics—to adjust the direct count numbers.

Second, extensive analysis by Census Bureau statisticians revealed that the ACE survey did not adequately measure a large number of double-counted persons and other counting mistakes

that are collectively called “erroneous enumerations.” The Census Bureau reported in October 2001 that the ACE survey had *overstated* the undercount by about 3 million people.

These two uncertainties led the Census Bureau to decide against relying on the ACE survey results to adjust for undercount in the 2000 Census. It should be emphasized, however, that the decision was based on the merits of using the ACE survey data to adjust the census number. Both demographic analysis and the ACE survey reported that there were net undercounts in the 2000 Census and that undercounts were higher for children and minorities. The Census Bureau’s decision was based on the technical merits of carrying out an adjustment and on their conclusion that the ACE results were less accurate than the actual enumeration, not on doubt about the existence of an undercount.

Advantages and Limitations of the ACE

The extensive research on the measurement and adjustment of census undercount, augmented with improved survey techniques for use in the 2000 ACE, promised a 2000 Census with improved accuracy. It is important, however, to be aware that census accuracy comes in many forms.

At the national level, the ACE was originally expected to move the national population count closer to the true population. The ACE was also expected to move the population counts for major race groups and for population subgroups by age and sex closer to the true population numbers. These efforts were designed to improve the distributive accuracy of the 2000 Census by providing more accurate population counts for groups who have historically been undercounted. The differential undercount between the white and black populations, for example, would be reduced to small differences in an adjusted 2000 count.

Demographers and statisticians debate the accuracy of adjusted data for smaller geographic areas. Whatever the coverage correction factors used in the ACE, the adjustments become much coarser as geographic detail becomes smaller. The accuracy for *some* small areas would be improved by adjustment; however, the Census Bureau reported that analysis "for counties with population below 100,000 indicated that the unadjusted census was more accurate" than the adjusted data.[15] Adjustment was found to improve the count "for areas in which the majority of the population resided," but there were serious concerns about the quality of the ACE data for smaller geographic areas.

Census Bureau Decisions on Adjustment and Implications for Data Users

The Census Bureau faced two decisions about the use of adjusted census data. The first dealt with the use of adjusted data for the purpose of redistricting, and the second dealt with the use of adjusted data for nonpolitical purposes. In March 2001, the Census Bureau released a set of detailed reports that showed discrepancies between the ACE estimates and the undercount estimates provided by demographic analysis. Citing doubts about the accuracy of the ACE estimates, particularly for small areas, the Census Bureau recommended that states use unadjusted census numbers for redistricting purposes. In October 2001, the Census Bureau recommended that unadjusted data also be used for nonpolitical purposes. A series of additional reports released by the Census Bureau showed that the ACE did not adequately measure a significant number of double counts and, as a result, overstated the national undercount numbers by about 3 million people.

However, undercount remains in the 2000 Census, and it is not without consequences. Differential net undercount means that inequities may exist for the drawing of political districts,

allocation of funds, and provision of social services based on census data. Differential net undercount also affects social science and public policy analysis, and should be considered when results based on census data are interpreted. The decision not to adjust the 2000 Census simplifies comparisons between 1990 and 2000 Census data. However, because the undercount varies in recent censuses, there is a need for some caution in interpreting trends over time. Data users should also be aware that certain population groups, particularly children and minorities, are more likely to be missed in the census. Reducing the undercount of these groups, in order to gain a complete and accurate count of the U.S. population, will be one of the major challenges faced by the Census Bureau in the in the next census.

Appendix A: How the ACE Worked in the 2000 Census

For the 2000 Census, the Accuracy and Coverage Evaluation (ACE) survey was conducted to gauge the quality of the 2000 Census enumeration and to produce independent estimates that could be used to adjust the census—provided that the ACE figures were judged to be more accurate than those of the census.

The ACE procedures in the 2000 Census involved matching information from an independent survey with census records from the initial direct enumeration. This process involved a separate survey operation—in other words, a survey that was independent and that had different staff and survey operations from the census enumeration—that conducted interviews in selected areas of the country. The results of the independent ACE survey were "matched," or compared, with census results to develop "coverage correction factors" for various population groups—factors that were then applied to the direct enumeration counts to provide a second set of census counts. The second set of counts was then analyzed to see if the counts were in fact more accurate than the unadjusted totals.

Sample Design. The ACE for the 2000 Census sampled almost 12,000 block clusters in the United States. The selected blocks were designed to give a representative sample of the U.S. population by race/ethnic origin, housing tenure (owner/renter), and other variables. The ACE sample included about 25 to 30 housing units in each block, yielding a final sample of about 314,000 housing units. This sample size was intended to be large enough to make fairly precise estimates of the undercount for specific population subgroups (i.e., groupings of the population by age, sex, race/ethnicity, tenure, and metropolitan status).

The independence of the ACE survey was critical to ensuring that the probability of a person being selected for the ACE was not affected by the probability of being selected in the census

direct enumeration. For this reason, the ACE hired its own staff and trained them separately from other 2000 Census operations. The ACE staff gathered their information through telephone interviews (for households that had already submitted their census questionnaire) and by visiting households door-to-door. The ACE interviewer attempted to conduct the interview with a household member—although not necessarily with the same household member who completed the initial census questionnaire. If the ACE interviewer could not conduct an interview with a household member, a proxy respondent was interviewed.

Matching. After the ACE independent interviewing was completed, the ACE and initial census records were matched through a process of automated computer matching and clerical review. This process allowed the Census Bureau to see how many persons might have been missed in the direct enumeration—and what characteristics they had. The matching could have indicated that an entire housing unit was missed, that a person within a household was missed, or that someone was erroneously included twice.

Based on previous field tests, the 2000 ACE operations were designed to minimize matching errors. It was important that the matching be as accurate as possible, because matching errors would have led to less precise estimates of census undercount in subsequent statistical analysis. For this reason, the Census Bureau designed procedures to avoid errors caused by incomplete, inaccurate, or misleading interview data. If needed, ACE staff conducted personal follow-up interviews to obtain additional information for accurate record matching.

ACE Adjustment for Undercount in Census Data. The Accuracy and Coverage Evaluation survey was based on a statistical technique known as “dual system estimation” or DSE. Statisticians at the Census Bureau and other survey research institutes have conducted considerable research into the uses of DSE. Their research has revealed that it is possible to

estimate the likelihood that people of varying characteristics (such as age, sex, race, urban/rural residence) will be included in the census enumeration. Because the ACE survey divided the nation's population into groups or strata who will have a similar probability of inclusion in the initial census, the separate ACE enumeration provided what is called a dual system estimate of the "true" population in each stratum.

The dual system estimates were then used to derive a coverage adjustment factor for each stratum to be applied to the original census data. The process occurred as follows:

- The nation's population was divided into hundreds of groups of people with similar characteristics, including race, sex, age, owner/renter status, urban/rural residence, and mail return rates. For example, one group might have included Hispanic white males, ages 18 to 29, in nonowner (renter) housing units, in metropolitan areas of 500,000 or more population, in census tracts with low mail return rates.
- The ACE was used to obtain a dual system estimate of the "true" number of people in each stratum. Analysts compared this alternate population with the initial census count; the ratio of the comparison is the coverage correction factor. For example, suppose the calculated coverage correction factor for a given stratum is 1.06. A factor of 1.06 means that there should be 106 person records for every 100 persons with similar characteristics in an area. If a census tract in Los Angeles, for example, enumerated 100 people with such characteristics, then an additional six people could be added for that particular tract.

The process described above was used to add people to the original census enumeration for small areas of the nation's population. The adjusted set of census data included people counted in the

initial enumeration and the persons added based on DSE estimates of their numbers and characteristics.

An Example of Dual System Estimation. Table 3 shows five blocks, each of which includes only one house. The "true" population for these five blocks is shown in the top panel. There are two people in block 1, three in block 2, two in block 3, one in block 4, and four in block 5. There are a total of 12 people in the true population for these hypothetical five blocks, although the number is initially unknown to the Census Bureau.

Suppose that the direct enumeration, the first census count, misses one person in block 3 and one person in block 5, producing an initial census count of 10 people. For this illustration, the possibility of missing an entire housing unit is not included, although it is certainly a possibility in an actual census.

Table 3. How the U.S. Census Bureau's Accuracy and Coverage Evaluation (ACE) Dual System Estimation Could Be Used to Adjust the 2000 Census

	Block number				
	1	2	3	4	5
True population (12 people)	2	3	2	1	4
Direct enumeration (10 people)	2	3	1	1	3
ACE count (6 people in 2 blocks)		2			4
Number of matches		2			3

Final adjusted census count =

$$\frac{\text{Direct enumeration} \times \text{ACE count}}{\text{Number of matches}} = \frac{10 \times 6}{2+3} = \frac{60}{5} = 12$$

Note: Based on an example in Wright, *American Scientist* (1998).

If the census count were not adjusted for undercount, the count of 10 people would be used for census reports and tabulations. In this simple case, the observed census undercount is 2

divided by 12, or 17 percent undercoverage rate. Next, the ACE takes a sample of blocks and conducts an independent census operation. Suppose the ACE selects blocks 2 and 5 for this operation. The ACE goes to the selected blocks and enumerates everyone in the housing units. In this example, the ACE misses one person in block 2, but achieves a complete count in block 5. The count for the ACE for the two selected blocks is six people.

The dual-system estimate for the final census count, adjusting for undercoverage in the initial direct enumeration is:

$$\text{Final census count} = \frac{(\text{direct enumeration count}) \times (\text{ACE count})}{\text{Number of matches}}$$

In the two blocks selected for the ACE, there are two matches in block 2 (person 1 and person 2 are counted in both the direct enumeration and the ACE counts) and three matches in block 5 (persons 1, 3, and 4 are included in both counts). This adds to a total of five matches.
For this example, the final census count is (10x6)/5 = 12. The dual system estimate in this case produces a final census count that is, in fact, the true population. In practice, the final count produced by the dual-system estimate might have resulted in fewer people (10 or 11) or more people (13 or 14). But, theoretically, the estimate would center on 12 people on average.

Acknowledgments

A number of people deserve special thanks for facilitating the completion of this report. My thanks to Census Bureau staff members Howard Hogan and Louisa Miller, who provided information about the 2000 Census process and helpfully answered my questions about specific procedures. I also offer my appreciation to others—John Haaga and Mark Mather, Population Reference Bureau; Sharon Lee, Portland State University's department of sociology; and William O'Hare, Annie E. Casey Foundation—who commented on drafts of this report.

Suggested Readings on the Census

Anderson, Margo J., and Stephen E. Fienberg. *Who Counts? The Politics of Census Taking in Contemporary America.* New York: Russell Sage Foundation, 1999.

Edmonston, Barry. "The 2000 Census Challenge." *PRB Reports on America* 1, no. 1, Population Reference Bureau, 1999.

Edmonston, Barry, and Charles Schultze, eds. *Modernizing the U.S. Census*. Washington, DC: National Academy Press, 1995.

Hoffman, Matthew. "Non-Census." *The New Republic* (February 9, 1998).

Kestenbaum, K. "Census 2000: Where Science and Politics Count Equally." *Science* 279, no. 5352 (1998): 798.

Lee, Sharon M. "Using the New Racial Categories in the 2000 Census," Washington, DC: A KIDSCOUNT/PRB Report, The Anne E. Casey Foundation and Population Reference Bureau, 2001.

Prewitt, Kenneth. "The U.S. Decennial Census: Political Questions, Scientific Answers." *Population and Development Review* 26, no. 1 (2000): 1-16.

Robinson, J. Gregory. "ESCAP II: Demographic Analysis Results," Executive Steering Committee for A.C.E. Policy II, Report No. 1. Washington, DC: U.S. Census Bureau, October 13, 2001.

Rousch, W. "Designing a Census in Which All Americans Count." *Science* 274, no. 5288 (1998): 713.

U.S. Census Bureau. "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy," accessed online at www.census.gov/dmd/www/pdf/Escap2.pdf.

U.S. Census Bureau. "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses," accessed online at www.census.gov/dmd/www/pdf/Recommend2.pdf.

U.S. Census Bureau. *Creating a Census for the 21st Century: The Plan for the Census 2000*. Washington, DC: U.S. Census Bureau, February 1997.

Wright, Tommy. "Sampling and Census 2000: The Concepts," *American Scientist* (May-June 1998): 245–253.

Endnotes

[1] Kirsten K. West and J. Gregory Robinson, "What Do We Know About the Undercount of Children?" accessed online at www.census.gov/population/www/documentation/twps0039/twps0039.html.

[2] For more information about the March 2001 decision, see the Census Bureau's website at www.census.gov/dmd/www/EscapRep.html.

[3] For more information about the October 2001 decision, sees the Census Bureau's website at www.census.gov/dmd/www/EscapRep2.html

[4] The Census Bureau's 1990 Post-Enumeration Survey (PES) indicated a net undercount rate of 1.6 percent in the 1990 Census, which is slightly lower than the rate indicated through demographic analysis (1.8 percent).

[5] Robinson, J. Gregory, "ESCAP II: Demographic Analysis Results," accessed online at www.census.gov/dmd/www/pdf/Report1.pdf

[6] U.S. Census Bureau, "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy," accessed online at www.census.gov/dmd/www/pdf/Escap2.pdf.

[7] Manuel de la Puente, "Using Ethnography to Explain Why People are Missed or Erroneously Included by the Census: Evidence from Small Area Ethnographic Studies," Center for Survey Methods Research U.S. Census Bureau, Working Paper (Washington, DC: U.S. Census Bureau, 1995).

[8] U.S. General Accounting Office, *Formula Grants: Effects of Adjusted Population Counts on Federal Funding to States*, GAO/HEHS-99-69 (Washington, DC: U.S. General Accounting Office, February, 1999).

[9] Barry Edmonston,. "The Case for Modernizing the U.S. Census," *Society* 39 (November-December 2001): 42–53.

[10] Ibid.

[11] Daniel O. Price, "A Check on Underenumeration in the 1940 Census" *American Sociological Review* 12 (1947): 44–49.

[12] Ansley J. Coale, "The Population of the United States in 1950 Classified by Age, Sex, and Color—A Revision of Census Figures," *Journal of the American Statistical Association* 50 (1955): 16–54.

[13] J.S. Siegel, "Completeness of Coverage of the Nonwhite Population in the 1960 Census and Current Estimates, and Some Implications," in *Social Statistics and the City: Report of a Conference Held in Washington, DC, June 23-23, 1967*, ed. D.M.Heer. (Cambridge, MA: Joint

Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, 1968).

[14] National Research Council, *Counting the People in 1980: An Appraisal of Census Plans*. (Washington, DC: National Academy Press, 1978).

[15] U.S. Census Bureau, "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy," accessed online at www.census.gov/dmd/www/pdf/Escap2.pdf.

The Annie E. Casey Foundation
701 Saint Paul Street
Baltimore, MD 21202
www.aecf.org

Population Reference Bureau
1875 Connecticut Avenue, NW, Suite 520
Washington, DC 20009
www.prb.org