CENSUS UNDERCOUNT AND THE QUALITY OF HEALTH DATA FOR RACIAL AND ETHNIC POPULATIONS

The United States was the first country in the world to undertake a regular census of its population. From its inception, the enumeration of nonwhite persons in the census has been problematic. Article I of the United States Constitution initially provided that American Indians who did not pay taxes would not be enumerated and that black slaves would be counted as three-fifths of a free person. After the three-fifths rule was abandoned, the Census Bureau continued to experience difficulty in counting the black population. In the 1870 Census, the undercount of the black population was estimated as in excess of 10%.1

In recent years, the US Census Bureau has devoted considerable resources to evaluating the extent of the undercount. The primary method used has been demographic analysis, which develops estimates of the population by combining various sources of administrative and demographic data. These analyses show that the census undercount for the total population declined steadily from 5.4% in 1940 to 1.2% in 1980, but increased to 1.8% in 1990.2 Racial and ethnic minority populations, the poor, and highly mobile persons have higher rates of undercount than the general population. The undercount for blacks, for example, decreased between 1940 and 1980 from 8.4% to 4.5%, but increased to 5.7% in 1990. In addition to demographic analysis, the 1990 Census allows for undercount to be estimated based on case-by-case matching of census records with those obtained in a postcensus enumeration survey of 165,000 households, conducted in the fall of 1990. According to the postcensus enumeration survey, the 1990 Census missed 0.7% of non-Hispanic whites, 4.6% of blacks, 5.0% of Hispanics, 2.4% of Asian and Pacific Islanders, and 12.2% of reservation Indians.3

Census undercount is a politically charged issue. The enumeration of the population is used to allocate seats in the House of Representatives to the various states, to establish congressional districts, and to distribute federal funds to state and local governments.

Census undercount may also distort our description of minority populations. Consider, for example, the impact of census undercount on gender ratios observed in the black population. According to the 1980 Census, within the 20- to 34-year age group, there were only 88 black men for every 100 black women. When these data are corrected for the next census undercount, there were 97 black men for every 100 women in this age group.4

The number of female-headed households may also be distorted by census undercount. According to the 1980 census, 40% of black households were headed by women. One small ethnographic study, conducted at the time of the 1970 Census, found that only 12% of officially classified "female-headed" households did not have a stable male cohabitant.5 In this study, adult male cohabitants who were present in a majority of black female-headed households and performed traditional domestic responsibilities were omitted in the enumeration of those households, because their presence was deliberately hidden from census enumerators.

Census undercount problems are of two types. First, there is the omission of entire households, due to the failure to enumerate a housing unit or the omission of a housing unit in an enumerated building. Second, there is the omission of some members of a household while others are enumerated. Most of the undercount of blacks is due to
within-household omission, particularly the omission of black males within households.

Census data play an important role in the development of health data. They are routinely used as denominators for the calculation of rates of specific health events, such as mortality rates. They are also used to construct sampling frames and adjust for nonresponse in population-based epidemiologic studies. Population undercounts can have important implications for our knowledge of the health status of the population. In particular, census undercount can inflate the rates of selected health events.

It is a mathematical axiom that any rate that uses an undercounted denominator is overestimated in exact proportion to the undercount in the denominator. Thus, all morbidity and mortality rates that use census data as denominators are overestimated by the same percentage as the undercount rate of the population in the denominator. Since the census undercount of African Americans is about 5%, vital rates for the total population, the total black population, and the overall black-white contrast are only minimally affected by the undercount. However, given that the undercount varies dramatically by age and gender, it can importantly inflate the rates of selected health events for particular subgroups. In 1990, the undercount for black males (8.5%) was more than twice that of black females (3%).

The National Center for Health Statistics (NCHS) does not adjust for undercount, but at least some NCHS reports discuss the undercount problem in technical appendices, and provide data for the adjustment of rates. The appendix of the report of the national mortality statistics for 1988 reveals that for all 5-year age groups of black men between the ages of 30 and 54 years, the undercount varies from 14% to 19%. Thus, all of the official mortality rates for most middle-aged black men are at least 14% too high! NCHS (1991) also indicates that adjustment for undercount would reduce the black-white ratio of deaths for homicide and legal intervention for males aged 35 to 39 years in 1980 from the officially reported 7.3 to 6.2.

Other health researchers have noted for some time that census undercount affects mortality rates for the black population. Rives reported that without adjustment for the net census undercount the mean expectation of life was underestimated by about 5% for blacks overall but by at least 15% for black women over age 70. Similarly, Kitagawa and Hauser indicated that between the ages of 20 and 54 years, mortality rates for black men were overestimated by 15% because of census undercount and age misstatement. More recently, it has been suggested that the current rates of AIDS infection among black males are biased upward because of the census undercount of that group.

Health researchers must remember that estimates of undercount are estimates. Demographic analysis estimates depend upon particular assumptions about errors in the census, as well as omission rates of births and deaths from the registration system, and immigration and emigration. Estimates from demographic analyses are very dependent upon the size of the undocumented immigrant population, the algorithms used to estimate racial-ethnic identification. A report of an American Statistical Association panel concluded that demographic analysis methods have become more unreliable in recent years, because of the influx of illegal aliens into the United States. Estimating the undercount by using a postcensus enumeration survey is also problematic because the survey is likely to miss the same people missed by the census.

Health researchers seldom consider the impact that census undercount may have on their reported distributions of disease. Demographic analyses reveal that black men have the highest rate of undercount of any race-sex age group. Since these estimates of undercount are available only at the nation-
al level, undercount rates may be considerably higher in some geographic areas. Despite the error in the estimates of undercount, adjusted data are probably closer to the truth than unadjusted data. In pursuit of the best available data, researchers should use adjusted data whenever available and should encourage the federal data system to do likewise.

REFERENCES

OMB DIRECTIVE 15: STANDARDS FOR THE CODING OF RACE AND ETHNICITY IN FEDERAL STATISTICS

Current standards for the coding of race and ethnicity in the federal government are given in Directive 15, issued by the Office of Management and Budget in 1977. The directive was intended to set minimum standards for the collection of information across agencies, to avoid duplication, and to include data on Hispanic populations, as required by Congress. The preface to Directive 15 noted, "These classifications should not be interpreted as being scientific or anthropological in nature. . . ." However, information collected under the directive has been used extensively not only to assess adherence to federal antidiscrimination legislation, but also for a wide variety of scientific and public health surveys and analyses. Congressman Thomas Sawyer, Chair of the Congressional Subcommittee on Census, Statistics, and Postal Personnel, concludes, "They become categories of convenience that may not serve the purpose for which they were originally put in place. As a result, we create an illusion of specificity, an illusion of precision where it may not exist."

Directive 15 implicitly uses four criteria to distinguish five populations (ie, American Indian or Alaskan Native, Asian or Pacific Islander, black, Hispanic, and white): (1) descent from the "original peoples" of a specified region, (2) specified cultural origins, (3) cultural identification or affiliation, and (4) race itself. However, not all criteria are used to define all populations. While the directive indicates that Hispanic identity is independent of race, it also provides a form of classification that excludes the possibility of Hispanic American Indian or Alaskan Native or Hispanic Asian or Pacific Islander. In addition, the directive is not clear about how the information on a person's race and ethnicity is to be collected, eg, by an interviewer observation or by questioning the individuals themselves. There is evidence that individuals may be ascribed differing race and ethnicity in different record systems.

Responding in part to growing racial-ethnic diversity in the United States, in part to