



A One-Number Census: Some Related History

Tommy Wright

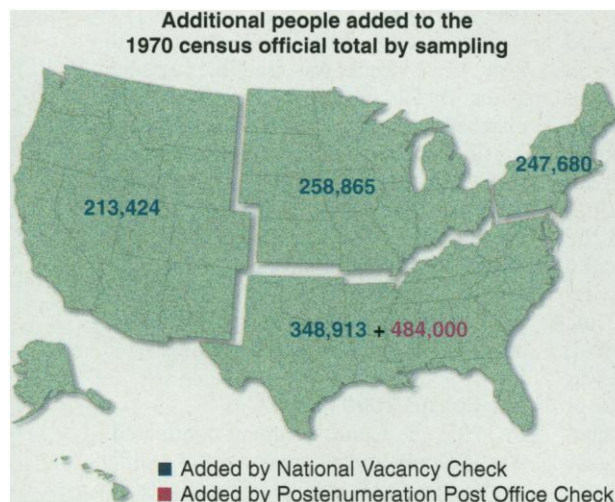
The U.S. Census Bureau plans to produce one best set of official counts of the population of the United States in the year 2000—a one-number census—by integrating the results of conventional counting techniques with results from probability sampling techniques. The plan will help lead to a result that includes more of the overall population, especially for certain subpopulations, and it will help control costs.

But not all favor such an approach (1). Congress has expressed concern about the constitutionality of sampling, the possibility that the use of sampling and estimation would allow the data to be manipulated for political advantage, and the magnitude of sampling error in very small geographical areas such as block levels. An agreement reached on 26 November 1997 between the White House and Congress directed the Census Bureau to pursue two plans: its current one-number census plan and a plan using only conventional counting techniques (without sampling). Carrying through on another part of that agreement, the U.S. House of Representatives sued to stop the one-number census; and a special three-judge federal court ruled on 24 August 1998 in its favor. The ruling and a similar one by a second federal district court are now being considered by the U.S. Supreme Court. Awaiting the Supreme Court ruling, it is instructive to reflect briefly on the need for and origins of the one-number census concept.

Errors from the Beginning

Since 1790, the U.S. Constitution has required a decennial census of the United States. The noble but elusive goal is to provide the true population of the nation on one specific date according to geography on levels as small as blocks and as large as states. The primary use of the census is to distribute seats in the U.S. House of Representatives among the states according to

their population sizes. Interest in the census heightened after Supreme Court decisions in the 1960s and 1970s that strengthened the “one person, one vote” concept, as well as after the passage of the Federal Revenue Sharing Act in 1972, under which sums of money began to be allocated to cities and states on the basis of the census. Each recent decennial census has seen the introduction of operations to improve this set of counts (2). The census data are also used for a wide variety of research purposes.



1970 Census. Sampling added ~1,552,882 people to the total.

Problems in obtaining an accurate count to the census appeared from the beginning (3). In reporting the 1790 results (officially a population of slightly under four million) to President Washington, Thomas Jefferson stated, “I enclose you a copy of our census, which so far as it is written in black ink, is founded on actual returns, what is in red ink being conjectured, but very near the truth.... Making very small allowance for omissions, which we know to have been very great, we may safely say we are above four millions.” In the 1830 census, both the original count and a corrected count were printed. Only a corrected count was printed in 1840. In the 1850 census, California state census tables from 1852 were used to estimate some 1850 California returns that were burned and missing. In the 1870 census, thousands of former slaves that were wandering the South after the Civil War

were omitted. After the 1920 census, disturbed by the results that first found more people living in urban than in rural places, Congress decided not to use the result for reapportionment.

Net Undercounting

By using records and estimation of births, deaths, immigration, and emigration, demographers have derived independent counts of population at the national level. Since 1940, the conventional count has been consistently smaller than the count from demographic analysis. Investigations suggest that the magnitude of undercounting exceeds the magnitude of overcounting, hence the phrase “net undercounting” (4, 5). Historical demographic analysis estimates of percent net undercount since 1940 are as follows (4): 1940 (5.4%), 1950 (4.1%), 1960 (3.1%), 1970 (2.7%), 1980 (1.2%), and 1990 (1.8%).

In 1990, the conventional census count (248,709,873) (6) also provided counts by various person types (for example, by race, age, ethnicity, and so on) for approximately 7 million blocks and for all levels of geography. Also in 1990, the Census Bureau combined the results from a nationwide probability sample (which provided reliable estimates of people missed and of people incorrectly enumerated) with the results from conventional counting to provide a second set of counts (252,712,821) (5) in similar detail (7). Several evaluations suggested that the second set of counts was superior to the first set at important levels (national, state, and large substate) of geography (8). However, the first set was used to apportion the U.S. House of Representatives. Just as a person with two watches never knows what time it is, a nation with two sets of numbers from a census would be torn, especially if the set believed to be superior cannot be used for certain applications. The debate over which set of numbers to use for 1990 ended on 20 March 1996 when the Supreme Court unanimously confirmed that Congress has the authority to conduct decennial censuses and that the Congress had delegated that authority to the Secretary of Commerce who favored the first set. Through research, the Census Bureau has been led to the approach of providing a one-number census in 2000.

The One-Number Census Concept

In a one-number census (9), the best possible single set of results based on counting, assignment, and estimation is used to pro-

The author is Chief, Statistical Research Division, U.S. Bureau of the Census, Washington, DC 20233, USA. E-mail: twright@census.gov

duce the most accurate census. "Counting" means the full array of techniques by which direct contact is made with all respondents (mail, personal visit, telephone, or other means), and it also means data obtained by proxy for another household, housing unit, or person. Historically, people have been added to the census by obtaining information about their existence from administrative records and verifying this information. "Assignment" is the use of indirect evidence from administrative records to add people to a specific geographic location without field verification. For Census 2000, research has convinced the Census Bureau that administrative records do not currently exist that would be adequate to reliably account for people without field verification, and assignment is not planned. "Estimation" is the application of statistical techniques (such as sampling) to account for people or units not directly counted or assigned.

Sampling at the Census Bureau

The U.S. Bureau of the Census has argued for sampling in the past (10), including in the mid-1930s, in which an Enumerative Check (Sample) was used to supplement and help control the quality of the nationwide voluntary Census of Unemployment during the Great Depression. "The Enumerative Check (Sample) achieved the recognition, in the Census Bureau and elsewhere, that large-scale sample surveys could make substantial contributions, and under appropriate design and control, could produce timely information that was more accurate than complete censuses or national registrations" (10). Greater accuracy tends to come with sampling because the improvements in measurement are greater than the loss of precision due to sampling.

In an effort to control and limit the extent of efforts to obtain needed information on every person captured in the 1940 census, sampling was introduced (11). These changes partly reflected the demand from government and the public for additional information for use in research and policymaking regarding unemployment, occupational shifts, migration, population growth, and so forth (11). In order to provide this data without requiring it of everyone, a sample of 1 out of 20 people nationwide was selected to answer supplementary questions. Although statistical estimates relating to the supplementary questions were made for the entire population, the population count was the result of summing the individuals captured on all of the collection forms nationwide (without the use of sampling). In the 1990 census, analogous supplementary data were collected from one out of every six housing units.

The 1970 Census

The 1970 census was the first census to be conducted in most areas by mail; it was also one that used two sampling efforts to contribute to the official census totals. The problems were (i) that the Census Bureau had found in pretests that occupied units incorrectly reported as vacant were a significant factor in the population undercounts (12) and (ii) that, from the 1960 census, housing unit coverage in the South was considerably worse than in the rest of the United States.

The first sampling effort, called the National Vacancy Check, selected for visits and interviews a sample of 13,546 housing units from a list of units that had been classified as vacant. Based on the sample results, approximately 8.5% of all the units initially classified as vacant were reclassified as occupied and an estimated 1,068,882 people—0.5% of the total 1970 census count—were added to the count (see the figure).

The second effort, the Postenumeration Post Office Check, was used in 16 southern states. In this check, the U.S. Post Office matched its list of addresses for certain areas (those counted by visits rather than mail) with the addresses from the census. From all addresses on the Post Office list but not on the census list, the Census Bureau selected a sample for visits. On the basis of the sample results, about 484,000 people were added, or 0.8% of the entire South and 0.2% of the total 1970 U.S. population (see figure).

The Census 2000 Plan

In 1995, the Census Bureau conducted a test of a one-number census at three different locations (13). At one of these sites, Paterson, New Jersey, conventional counting with a follow-up sample of nonresponding housing units yielded a count of 130,832 people. When these results were combined with the results of a quality check using sampling, the resulting one-number census for Paterson was 148,394 people (14).

The Census Bureau is just concluding a dress rehearsal at three other diverse sites to demonstrate that its Census 2000 Plan (15–18) is not only theoretically sound but operationally feasible. It represents the coming together of many operations and activities that were individually designed and tested on earlier occasions. The big question in such a dress rehearsal is "Do the pieces fit together well in accomplishing the goal?" It also allows the Census Bureau yet another opportunity for refinement. Consistent with a compromise between Congress and the White House, two of the sites received a one-number census treatment whereas the third received only a conventional counting treatment. Results are under review.

The United States is currently at a crossroads. The evidence, repeatedly from the first census through the 1990 census, suggests that if Census 2000 were conducted on different occasions using the conventional methods of the past, even with increased outreach, the results would tend to be consistently below the truth. On the other hand, theory, simulations, and tests lead us to believe that a one-number census conducted on different occasions would tend to yield results around and closer to the truth.

In addition to being accurate and of consistent quality across states and local areas, it is absolutely essential that the United States' ultimate plan for the census in 2000 have widespread support. If people will be missed or incorrectly enumerated in 2000, there is no guarantee that they will live in the same places or have the same characteristics of past missed and incorrectly enumerated groups. The one-number census approach offers protection against these variations.

References and Notes

1. D. A. Freedman, *Science* **252**, 1233 (1991); M. Eaton et al., *SIAM News* **31** (9), 10 (1998).
2. E. D. Goldfield, *Innovations in the Decennial Census of Population and Housing: 1940–1990* (Commissioned Paper Prepared for The Year 2000 Census Panel Studies, Committee on National Statistics, National Research Council, Washington, DC, 1992).
3. M. D. Rosenthal, *Striving for Perfection: The History of Undercount in the Census* (draft unpublished paper, U.S. Bureau of the Census, Washington, DC, 1998).
4. J. G. Robinson, B. Ahmed, P. Das Gupta, A. Woodrow, *J. Am. Stat. Assoc.* **88**, 1061 (1993).
5. *Assessment of Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates* (Attachment 3 to the Report of the Committee on Adjustment of Postcensal Estimates, U.S. Bureau of the Census, Washington, DC, 7 August 1992).
6. *1990 Census of Population: General Population Characteristics, United States (1990 CP-1-1)* (U.S. Bureau of the Census, Washington, DC, 1990), table 3, p. 3.
7. H. Hogan, *Am. Stat.* **46**, 261 (1992).
8. See, for example, the special section of *J. Am. Stat. Assoc.* **88**, 1044 (1993).
9. S. M. Miskura, *Definition, Clarification, and Issues: "One-Number Census"* (memorandum, U.S. Bureau of the Census, Washington, DC, 14 April 1993).
10. M. H. Hansen and W. G. Madow, in *On the History of Statistics and Probability*, D. B. Owen, Ed. (Dekker, New York, 1976), pp. 75–102.
11. F. F. Stephan, W. E. Deming, M. H. Hansen, *J. Am. Stat. Assoc.* **35**, 615 (1940).
12. *Effect of Special Procedures to Improve Coverage in the 1970 Census* [Report PHC(E)-6 in the Evaluation and Research Program of the 1970 Census of Population and Housing, U.S. Bureau of the Census, Washington, DC, 1974].
13. E. A. Vacca, M. Mulry, R. A. Killion, *The 1995 Census Test: A Compilation of Results and Decisions* (Memorandum 46, 1995 Census Test Results, U.S. Bureau of the Census, Washington, DC, 1996).
14. T. Wright, *Construction of a One-Number Census: An Illustration* (unpublished paper, U.S. Bureau of the Census, Washington, DC, 1998).
15. T. Wright, *Am. Sci.* **86**, 245 (1998).
16. J. E. Farber, R. E. Fay, E. L. Schindler, in preparation.
17. J. H. Thompson and R. E. Fay, *Proc. Am. Stat. Assoc.*, in press.
18. *Census 2000 Operational Plan, April 1998 (Revised)* (U.S. Bureau of the Census, Washington, DC, 1998).
19. This article reports the results of research and analysis undertaken by Census Bureau staff. The views expressed are the author's and do not represent those of the U.S. Census Bureau.