The 2010 Census: How It's Done and Why It Matters

Stanley K. Smith Professor, Department of Economics and Director, Bureau of Economic and Business Research University of Florida

Paper presented at the annual meeting of the Association for University Business and Economic Research, Austin, TX October 18-20, 2009 Where do you live? For many people this seemingly simple question doesn't have a simple answer. Some retirees spend winters in Florida or Arizona and summers in New York or Minnesota. Others buy an RV and move from place to place, with no fixed place of residence. College students spend part of the year in their college towns and part in their home towns. Migrant farm workers often move from place to place over the course of a year, spending no more than a few weeks or months at any given location. Children of divorced parents may shift between the mother's home and the father's home on a monthly, weekly, or even a daily basis. Where do these people live?

This is just one of the issues the U.S. Census Bureau will have to deal with in the upcoming census. The 2010 Census will attempt to count every person in the United States—regardless of their age, sex, race, citizenship, or even legal status—and assign them to a specific place of residence. This is a difficult task but is extremely important because the decennial census determines how many congressional representatives each state will have during the following decade; provides the raw material for establishing state and local voting districts; affects the distribution of more than \$300 billion in federal dollars each year; and forms the basis for developing a variety of socioeconomic and demographic statistics. The importance of the census cannot be overstated. It is the most comprehensive source of demographic data in the United States and tells us much of what we know about our nation's population and how it is changing over time.

In this paper, I will describe a bit of the history of census-taking in the United States, provide an overview of how the census is conducted, and discuss some of the hotbutton issues surrounding the 2010 census. I will confine my comments to the decennial

census of population and housing and will not cover other censuses or surveys conducted by the Census Bureau.

A Brief History of the Census

[Figure 1. Census Taker]

An enumeration of the population is required by Article I, Section 2 of the U.S. Constitution, but the manner in which this enumeration is carried out is left up to Congress. The first census was conducted in 1790 and was used as a basis for apportioning seats in Congress and levying taxes on states. Subsequent censuses have been conducted at 10-year intervals ever since.

The U.S. was certainly not the first nation to take a census. The earliest recorded census was undertaken around 3800 BC by the Babylonians (Skerry, 2000). Caesar Augustus conducted one 2000 years ago, as noted in the Christmas story in the Gospel of Luke. But the U.S. was the first country in history to take a census at regular intervals from its very beginning as a nation (Ashabranner and Ashabranner, 1989) and was the first to use a census to apportion political power (Anderson and Fienberg, 1999).

The first census was taken by U.S. marshals and their assistants, who traveled on horseback throughout the 13 original states and four districts or territories (U.S. Census Bureau, 2002). They collected information on the number of free white males age 16 or older, free white males less than age 16, females of any age, all other free persons, and slaves. Most American Indians ("Indians not taxed") were not counted at all. Although slaves were counted, they were given a weight of only 3/5 for apportionment purposes (the famous 3/5 compromise). The marshals were not trained as census enumerators and did not

follow any consistent procedures or even use a uniform questionnaire. It took them 18 months to visit all the households and tally the results (Anderson, 1988).

[Table 1. U.S. Population since 1790]

The 1790 census counted 3.9 million people. Both George Washington and Thomas Jefferson thought the count was about 200,000 too low. Washington blamed people who concealed information because they wanted to avoid taxes; he also thought many of the enumerators were lazy and didn't do their jobs properly (Ashabranner and Ashabranner, 1989). So there have been complaints about the quality of the census ever since the first one was conducted 220 years ago.

There were other controversies as well. The Constitution did not specify how census data should be used to apportion seats in Congress, and different plans led to different results. Thomas Jefferson and Alexander Hamilton each came up with a plan and, after much debate, Congress passed a bill based on Hamilton's plan. Many southerners objected to this plan, however, and George Washington used his first veto to quash it; Jefferson's plan eventually became law (Anderson, 1988). Clearly, politics have played a role in the collection and use of census data from the very beginning.

Many changes have been made over the years, both in terms of census procedures and the questions asked. These changes reflected changes in society and the need for statistical data, as well as changes in technology. More detailed information on age, sex, and race were collected as the decades went by, and questions were added regarding marital status, number of children, state of birth, income, education, occupation, and many other characteristics. Information on slave status was collected from 1790-1860, but not thereafter. In response to concerns about a rapidly growing immigrant population, the 1920

census added several questions on the foreign-born population. In response to changing living arrangements, the 2000 census added a question on grandparents as caregivers and dropped a question on childbearing. The 1890 census had the largest number of questions going to all households, either before or since (U.S. Census Bureau, 2002).

Significant improvements in census procedures were made in the latter part of the 19th century. A comprehensive set of maps was constructed in 1870, federal marshals were replaced by trained enumerators in 1880, and punch card tabulating machines were introduced in 1890. The Census Bureau became a permanent federal agency in 1902 (Anderson, 1988).

The reference data for the census has bounced around over time. The first census was taken in August, but others have been taken in June, April, and January. The date has been fixed as April 1 since 1930 (Skerry, 2000).

The 1940 census was the first to include a census of housing and the 1950 census was the last in which all data were collected by enumerators going door-to-door. In 1960, the Census Bureau sent questionnaires to about 60% of households by mail, asking respondents to fill out the forms and hold them until a census enumerator picked them up. This worked so well that the Census Bureau started using a mail out/mail back system for most households in 1970.

The modern census is largely a self-enumeration in which people count themselves. This puts a premium on public cooperation and requires that the Census Bureau develop clear questions and easy-to-follow instructions, mount a good advertising campaign, provide language assistance when needed, and convince people to participate.

The 2000 census collected information on age, race, ethnicity, marital status, education, income, occupation, place of birth, age of housing unit, number of rooms, plumbing facilities, and many other characteristics. It was the largest peacetime mobilization in the nation's history:

-- It employed as many as 860,000 workers, including some 500,000 temporary workers.

-- More than 400 million questionnaires were printed. This took two printers, running 24 hours a day, 7 days a week, for 10 months.

-- Mailings filled nearly 500 tractor-trailers.

Printing the forms took more than 14,000 tons of paper and 15,000 gallons of ink.
Four data capture centers processed 1.5 billion sheets of paper. If all these sheets had been stacked on top of each other, the stack would have been 100 miles high.

[Merrill, 2000; Merrick, 2001; Prewitt, 2003]

The 2010 Census

It's now 2009 and the next census is less than six months away. Conducting a complete count of the nation's population is a formidable task, given the huge number of people, the vast geographic area, and the tremendous educational, cultural, and linguistic diversity found in the United States. How does the Census Bureau do this?

Planning for the 2010 census began during the 1990s, but the starting point for the enumeration itself is a list of addresses based on the list used in 2000. This list has been updated using data on building permits, post office delivery records, tax records, and other administrative records. Earlier this year, thousands of enumerators went door to door, adding addresses that were missed and deleting addresses with no housing units. State and

local governments were also given the opportunity to review the list and note any discrepancies. Final updates will be made early next year.

Around the middle of March, 2010, census forms will be sent to the large majority of addresses on the list; for addresses without clearly defined house numbers and street names (mostly in rural areas), forms will be delivered by census enumerators. People will be asked to fill out the forms and mail them back. In a few sparsely populated or remote areas (e.g., parts of Alaska), census workers will visit the household and fill out the forms themselves. Special procedures will be followed for the homeless and residents of group quarters facilities such as prisons and college dormitories.

Reminder postcards will be sent out soon thereafter and if the forms are not returned within several weeks, census enumerators will follow up with telephone calls and personal visits. If they cannot find anyone after six phone calls or visits, they will attempt to collect information from a landlord or neighbor. If even that fails, occupancy status will be estimated based on the enumerator's judgment, and personal characteristics will be imputed based on the characteristics of a nearby household. So, even if a person refuses to participate or is somehow missed, he or she may still be included indirectly through this imputation process. In 2000, this type of imputation raised the final population count by 1.1 million persons (Prewitt, 2003).

All the census forms will be sent to one of three central processing centers (Baltimore, Phoenix, and Jeffersonville, Indiana) where they will be tabulated into counts for states, counties, cities, and other administrative and political units down to the block level. By law, population counts for states must be completed by December 31 of the

census year and counts for local areas by April 1 of the following year. Reports providing additional data will be released during the following months and years.

For the first 100 years, tabulations of census data were done strictly by hand. In 1890, a tabulating machine based on punch cards was introduced. This machine was invented by Herman Hollerith, a Census Bureau employee. Hollerith patented his machine and left the Census Bureau to form his own company, which eventually merged with several others to form a new company, IBM. The Census Bureau started using a UNIVAC computer system in 1950 and all records are now tabulated electronically. Although the Census Bureau still publishes a number of printed reports, data are available electronically and can be found on the Census Bureau website.

Prior to 1940, a single census form containing all questions was sent to every household. Starting in 1940, two forms were used in order to ease the burden on respondents and reduce the cost of collecting and tabulating data. A short form went to most households, collecting data on a limited number of characteristics (e.g., age, sex, race, and home ownership). A long form went to a sample of households, collecting additional data on income, education, occupation, age of housing unit, number of rooms, and so forth. Over the years, the sample has accounted for 15-20% of all households.

For 2010, the Census Bureau will go back to using a single short form for every household and will collect the more detailed data through the American Community Survey (ACS). The ACS is a sample survey begun on a trial basis in four test sites in 1996 (Mather, Rivers, and Jacobsen, 2005). Since that time it has grown to cover the entire nation, reaching some 3 million households each year. It collects data on a monthly basis throughout the year and is used to develop annual estimates of demographic and housing

characteristics for every state, county, city, census tract, and block group in the United States. For places with 65,000 or more residents, estimates are based on data collected in a single year. For smaller places, estimates are based on data collected over a 3-5 year period. In essence, the ACS has replaced the long form of the decennial census.

Accuracy of Census Counts

How accurate is the census? Coverage error refers to the extent to which the census counts all the people (Edmonston and Schultze, 1995). Net coverage error—the difference between the number of people missed and the number counted more than once—can be measured in two ways:

1) Demographic analysis – Historical data on births, deaths, and migration are combined with data from other sources (such as Medicare records) to form an estimate of what the population should be on April 1 of the census year. The difference between this estimate and the census count is considered to be coverage error. This technique works well at the national level, but not for states and local areas because of inadequate data, especially for migration.

2) Post-enumeration surveys – A sample of blocks is drawn and the population of those blocks is re-enumerated soon after the census is completed. Coverage error is estimated by comparing people included in only one of the two counts with those included in both. This is similar to the capture-recapture method used by wildlife biologists to estimate animal populations and can be used for coverage estimates of states and local areas.

[Table 2. Census Enumeration Errors 1940-2000]

Coverage error was first measured following the 1940 census, when it was estimated that there was a net undercount of 5.4%. Accuracy improved in every census through 1980,

when the net undercount was calculated as 1.2% (Edmondson and Schultze, 1995). In 1990, however, the net undercount increased to 1.6% (Citro, Cork, and Norwood, 2004). Why did undercount go up between 1980 and 1990?

The most obvious reason is that the mail return rate declined, from about 81% of households receiving census forms in 1980 to 75% in 1990 (Citro, Cork, and Norwood, 2004). Having to follow up on so many households made it more difficult to collect accurate data and raised the net undercount rate. But why did the mail return rate decline? The answer to this question is not as clear. It probably had to do with increased cultural and linguistic diversity (e.g., more people who didn't speak English well), more complicated living arrangements (e.g., shared custody of children, blended families), greater seasonal mobility (e.g., snowbirds), an increase in the number of undocumented immigrants, and increased cynicism and distrust for anything related to Big Government.

The Census Bureau received a lot of criticism regarding the 1990 census and made major efforts to improve the quality and reduce the cost of the 2000 census, mounting a professional advertising campaign, designing a more user-friendly questionnaire, working more closely with state and local officials, and using more up-to-date technology. It appears that those efforts were successful. The mail return rate rose from 75% to 78%, which lowered costs and allowed more time for follow-up with households that did not return their questionnaires.

Demographic analysis showed a net undercount of only 0.1% in 2000; the postenumeration survey actually showed a net *overcount* of about 0.5%. That is, more people were double-counted than were missed by the census. This doesn't mean there were no errors, however, because some places and demographic subgroups were undercounted while

others were overcounted. For example, blacks and Hispanics had a net undercount of 1.8% and 0.7%, respectively, while non-Hispanic whites had a net overcount of 1.1%. It should also be noted that although the net undercount rate was very small, there were an estimated 16 million omissions and 17 million erroneous inclusions, for a gross error of almost 12% (Citro, Cork, and Norwood, 2004).

Confidentiality of Census Data

In the early years, there were no restrictions regarding the confidentiality of census data. In fact, data for individual households were posted in public places so that people could check them for accuracy (U.S. Census Bureau, 2003). But as time went by, people started becoming concerned about the sensitive nature of the information collected. Informal restrictions were placed on the release of individual data in 1850. These restrictions were formalized in the Census Acts of 1880 and 1890. The 1910 Census Act did not prohibit disclosure, however, and individual data were subsequently used by the U.S. Department of Justice to prosecute individuals arrested for draft evasion during World War I. Data from the 1920 census were used to identify people with specific characteristics such as literacy and citizenship. This information was used for purposes such as targeting literacy campaigns and deporting illegal immigrants (U.S. Census Bureau, 2003).

Concerns about the disclosure of census data rose again during the 1920s and the Census Act of 1929 imposed strict confidentiality requirements. Subsequent requests for individual data—even by federal law enforcement agencies—were turned down. The only instance in which confidentiality may have been breached occurred in 1942, when the Census Bureau provided the War Department with information on the location of Japanese Americans. This information was used to identify thousands of people who were rounded

up and sent to internment camps. It is still not clear, however, whether the Census Bureau released the names and addresses of individual people or simply identified blocks and census tracts with large concentrations of Japanese Americans.

Strict confidentiality restrictions were formalized in 1954 in Title 13 of U.S. Code. Since that time, only sworn Census Bureau employees have had access to data in which individuals can be identified, and then only for purposes of aggregation and tabulation. In fact, aggregate statistics for specific groups are suppressed if there are not enough people in the group to maintain confidentiality for any respondent. Individual census data become public record after 72 years and are often used for genealogical purposes, as people try to find information on parents, grandparents, and so forth.

Issues and Controversies

As we look ahead to the 2010 census, what are the hot-button issues that will dominate the headlines? I will mention five that I think will be important: costs, privacy and confidentiality, racial classification, residence rules, and undercount adjustments.

[Table 3. Census Costs 1970-2000]

1) The cost of the decennial census has increased substantially over time. Although inflation and population growth certainly played a role, costs per person in real terms have increased as well. The inflation-adjusted cost per housing unit more than quadrupled between 1970 and 2000; it is expected to grow by another 60% in 2010, reaching more than \$14 billion.

These increases were caused partly by the difficulty of conducting a census in a nation that is increasingly diverse, mobile, and less trustful of government. But they were also caused by mistakes within the Census Bureau itself. There were several blunders

leading up to the 2010 census (e.g., problems with hand-held computers that led to expensive changes in follow-up procedures) and the Census Bureau has been slow to adapt to technological changes (e.g., use of the Internet). Although \$14 billion may seem like a drop in the bucket compared to overall government spending, it is still a substantial outlay. You will hear complaints about the cost of the census in the coming months and there will be pressure to reduce those costs in 2020.

2) In spite of legislation protecting the confidentiality of census returns, concerns about privacy and confidentiality have increased in recent decades. Some people believe the census is an invasion of privacy and the government has no right to collect personal information. Some radio talk show hosts and right-wing political groups have made this argument, urging their listeners and supporters not to participate in the census. Others are concerned that individual data will be released despite confidentiality requirements. Illegal immigrants—and others involved in illegal activities of one type or another—may be particularly sensitive on this issue.

Concerns about privacy and confidentiality form the philosophical basis for refusing to participate in the census. Although answering census questions is required by law (Title 13, U.S. Code), I don't know of anyone who has been prosecuted for failing to cooperate. If enough people refused to participate, it would reduce accuracy and raise the cost of the census. Perhaps more important, if there is any breach in the *confidentiality* of the 2010 census, it will be a public relations disaster and threaten future census-taking.

3) Another critical issue is how to define and measure race and ethnicity. Race and ethnicity are socially as well as biologically determined characteristics; consequently, they are difficult to measure—or even to define—especially in a self-enumeration. The choice of

the racial and ethnic groups to include in the decennial census—and the ways in which those groups are identified and measured—is a politically-charged issue because race and ethnicity have become central to public policy in recent years and census numbers are often used as a measure of political power (Skerry, 2000).

The census questionnaire asks people to classify themselves as Hispanic or non-Hispanic and as belonging to one or more of several racial groups, including white, black, American Indian or Alaska native, and a variety of Asian and Pacific Islander groups. These categories were chosen by the Office of Management and Budget, not by the Census Bureau. They have been strongly influenced by the input of various advocacy groups and have changed over time. Given the importance of racial and ethnic data for funding and other decision-making purposes, it is likely that there will be political pressure to make further changes in the future. Both the choice of racial and ethnic categories and the accuracy of the resulting data are likely to be hot-button issues in 2010.

4) Residence rules are also controversial. The Census Bureau not only has to count everyone residing in the United States, but also has to assign them to a specific geographic location. There are two ways this could be done. One is to count people in the place they are physically located on April 1, regardless of how long they have been there or how long they intend to stay (the *de facto* approach). The other is to count people at their usual place of residence, even if they are not physically there on April 1 (the *de jure* approach). The latter is the approach the U.S. has followed ever since the first census in 1790, but it doesn't always lead to clear outcomes. What is the usual place of residence for college students, snowbirds, migrant farm workers, and citizens living abroad? The Census Bureau has

developed general guidelines for answering this question, but a great deal of uncertainty and ambiguity remains.

Residence rules are important—and controversial—because they can have a major impact on population counts. One example of this controversy is the lawsuit filed by the State of Utah following the 2000 census (Prewitt, 2003). Except for members of the armed forces and federal government employees (and their families), the census does not include citizens living abroad. Utah argued that the Census Bureau should count *all* citizens living abroad, including some 11,000 Mormon missionaries the State claimed were Utah residents. If those people had been included, Utah would have gained another seat in Congress (at North Carolina's expense). The U.S. Supreme Court ruled against Utah but—partly as a result of this lawsuit—the Census Bureau tested the feasibility of counting U.S. citizens living abroad. This is a complex issue, both in terms of defining who should be included and figuring out how to count them. As of now, the Census Bureau has no plans to count citizens living abroad.

5) Finally—and perhaps most important—is whether census counts should be adjusted to account for likely errors. The fact that the census misses some people and double-counts others has led to calls for the Census Bureau to develop statistical adjustments to census counts. A number of states, counties, and cities have sued the federal government, petitioning for adjustments of one type or another. In one famous case, the City of Detroit filed suit in the spring of 1980—long before census results were known seeking an adjustment for what they claimed would be an "inevitable undercount" of racial minorities; in all, 54 suits related to the 1980 census were eventually filed (Anderson, 1988).

Similar suits were filed following the 1990 and 2000 censuses and there will probably be more following the 2010 census. To date, none have been successful.

The Census Bureau has tested a number of ways to adjust census counts and many statisticians and demographers believe it could be done in a way that would improve overall accuracy. However, such adjustments would not improve accuracy everywhere; in some places, in fact, they would undoubtedly *reduce* accuracy (especially places with small populations).

This is a controversial issue because adjustments would lead to winners and losers in terms of political representation and the distribution of funds. On average, blacks, Hispanics, renters, and low-income people tend to be undercounted, whereas whites, homeowners, and high-income people tend to be overcounted. Consequently, places with lots of the former typically favor adjustment and places with lots of the latter oppose it. Democrats generally favor adjustment because undercounted groups tend to vote Democratic; Republicans oppose it for the same reason.

The crux of the issue from a legal standpoint is what constitutes an "actual enumeration" of the population, as required by the Constitution. Critics claim statistical adjustment is not an enumeration and therefore violates the Constitution. Proponents claim statistical adjustment is simply part of the modern enumeration process. Scientific sampling and statistical adjustment were unknown when the Constitution was written. Would the authors have prohibited those procedures if they had known about them? Who knows?

The U.S. Supreme Court ruled in 1999 that adjusted counts could not be used for reapportionment following the 2000 census. However, it did not rule on whether statistical adjustment *itself* was unconstitutional; that issue is still unresolved. Also, the court did not

prohibit the use of adjusted counts for purposes such as drawing political boundaries and distributing government funds. The George Herbert Walker Bush administration decided not to adjust census counts following the 1990 census and the George W. Bush administration decided the same thing following the 2000 census. It remains to be seen what the Obama administration will do following the 2010 census. Census adjustment is an explosive (and emotional) political issue that you will hear much more about during the coming months and years.

Conclusion

The decennial census is a huge and complicated undertaking and, all in all, I believe the U.S. Census Bureau has done an excellent job carrying it out for many years. The census is vitally important because it has a major impact on the distribution of political power in the United States, it determines the allocation of hundreds of billions of dollars each year, and it tells us how we are changing as a nation. It is no wonder the decennial census is the source of so many controversies. Figuring out how to conduct it efficiently, accurately, and fairly will be a challenge not only in 2010, but in the decades to come.

REFERENCES

Anderson, Margo J. 1988. *The American Census: A Social History*. New Haven: Yale University Press.

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

Ashabranner, Melissa and Brent Ashabranner. 1989. *Counting America: The Story of the United States Census*. New York: G.P. Putnam's Sons.

Citro, Constance F., Daniel L. Cork and Janet L. Norwood (eds.). 2004. *The 2000 Census: Counting under Adversity*. Washington DC: National Academy Press.

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

Mather, Mark, Kerri L. Rivers, and Linda A. Jacobsen. "The American Community Survey," *Population Bulletin* 60, No. 3. Washington DC: Population Reference Bureau.

Merrick, Amy. 2001. "Counting on the Census." The Wall Street Journal, February 14.

Merrill, Dave. 2000. "Census 2000: Turning a Pile of Paper into the USA's Digital Portrait." *USA Today*, March 20.

Prewitt, Kenneth. 2003. *Politics and Science in Census Taking*. New York: Russell Sage Foundation.

Skerry, Peter. 2000. *Counting on the Census? Race, Group Identity, and the Evasion of Politics*. Washington DC: The Brookings Institution.

U.S. Census Bureau. 2002. *Measuring America: The Decennial Censuses from 1790 to 2000.* Washington DC: Government Printing Office.

U.S. Census Bureau. 2003. *Census Confidentiality and Privacy: 1790-2002*. Unpublished report.

U.S. Census Bureau, 2008. *Statistical Abstract of the United States: 2009.* Washington DC: Government Printing Office.



	Population	Change	%
Year	(millions)	(millions)	Change
1790	3.9		
1800	5.3	1.4	35.1
1810	7.2	1.9	36.4
1820	9.6	2.4	33.1
1830	12.9	3.3	33.5
1840	17.1	4.2	32.7
1850	23.2	6.1	35.9
1860	31.4	8.2	35.6
1870	39.8	8.4	26.6
1880	50.2	10.4	26.0
1890	63.0	12.8	25.5
1900	76.2	13.2	21.0
1910	92.2	16.0	21.0
1920	106.0	13.8	15.0
1930	123.2	17.2	16.2
1940	132.2	9.0	7.3
1950	151.3	19.1	14.5
1960	179.3	28.0	18.5
1970	203.3	24.0	13.4
1980	226.5	23.2	11.4
1990	248.7	22.2	9.8
2000	281.4	32.7	13.1
2010*	310.2	28.8	10.2

Table 1. US Population 1790-2010

* Estimated

Source: U.S. Census Bureau, 2008. *Statistical Abstract of the United States: 2009.* Washington DC: Government Printing Office.

Year	% Undercount	
1940	5.4	
1950	4.1	
1960	3.1	
1970	2.7	
1980	1.2	
1990	1.6	
2000	0.1	

Table 2. Census Undercount 1940-2000 (Demographic Analysis)

Sources: Barry Edmonston and Charles Schultze (eds.), *Modernizing the U.S. Census*. Washington DC: National Academy Press, 1995; Constance Citro, Daniel Cork and Janet Norwood (eds.), *The 2000 Census: Counting under Adversity*. Washington DC: National Academy Press, 2004.

Table 3. Census Costs 1970-2000

	Cost in 2000	
	Dollars	Cost per Unit
Year	(billions)	(2000 Dollars)
1970	0.9	13
1980	2.2	24
1990	3.3	32
2000	6.6	56

Source: Constance Citro, Daniel Cork and Janet Norwood (eds.), *The 2000 Census: Counting under Adversity*. Washington DC: National Academy Press, 2004..