Some Problems in Using Federal Census Data in Longitudinal Analyses of Old Age

The Underenumeration of the Aged in the 1870 and 1880 Occupational Statistics

One of the canons of historical analysis is that the credibility of an interpretation depends in large part on the reliability and quality of the data a researcher uses. Every bit of evidence must be examined for internal and contextual discrepancies. Very often a historian can cope with deficiencies he or she detects by making adjustments or estimates and by being sensitive to the ways in which biases in sources can distort the mathematical and theoretical interpretations of the analyses performed. Occasionally a researcher finds evidence so flawed, however, that it is useless.

Occupational statistics for men and women over sixty years old in the 1870 and 1880 federal censuses are a case in point. Although constituting the earliest age-specific statistics on the number of older people gainfully employed, the figures arouse suspicion: the percentages for men over sixty reported as engaged in all occupations in 1870 and 1880 were 64.2% and 64.3% respectively; in 1890, however, 73.2% of all men over sixty-five were listed as employed. These statistics seem curious because work force participation generally tends to decline with advancing age. In fact, on the basis of internal evidence in various federal documents as well as a comparison of federal and state census statistics on the work force participation rates of elderly men and women in Massachusetts and New York during the latter part of the nineteenth century, I determined that the 1870 and 1880 old-age data were grossly underenumerated.¹

Changes in instructions to enumerators taking the census in large measure may explain why census returns on the percentage of elderly men and women working at gainful jobs were less accurate before 1890 than after. While enumerators were told that inquiring about people’s profession,
occupation, or trade was "one of the most important questions" of the 1870 and 1880 schedules, there was no specific instruction requiring that occupational data be gathered. More importantly, no fines were levied if such data were not collected. This ambiguity was eliminated in 1890, when the census office required enumerators "to obtain each and every item of information and all the particulars." Information relating to age, sex, and race, on the other hand, had to be gathered for every person; no return would be accepted whenever these data were missing. As a result, the occupational status of every member of every household was recorded in 1890. This practice continued in subsequent years, so information on this item should be reasonably complete.

Other Census Data Problems

The problem of underenumeration, needless to say, is not limited to occupational statistics in the 1870 and 1880 censuses. For instance, canvassers often overlooked certain segments of the population (notably blacks, immigrants, and urban dwellers) in doing surveys. This means that the totals for various subsets of older Americans in a given census year, especially in the nineteenth century, might have been greater than reported. Often those interviewed tainted the reliability of information gathered. The number of persons who said that they were sixty years old or older probably differs from the number who actually were that precise age. A few people guessed because they did not know how old they really were. Other people lied about their age; the reasons motivating them were so diverse, however, that it is impossible to determine the net direction of the distortion. Age heaping—resulting from people's tendency to round off their ages to the nearest five-year interval or decade—compounds the problem.

In addition, systematic alterations in the way that the categories were defined and the survey itself was conducted make it difficult to trace changes over time. Before the Civil War, census officials classified all slaves and free blacks over forty-five together as one age group, thereby preventing comparison of the aged by race in the antebellum period. Analyses of aged blacks are further hampered by a bureaucratic decision to classify all blacks, Indians, and Orientals as "colored" in 1880 and 1890. (Although it is reasonable to suppose that nearly all "colored" people in the South at the time were blacks, such an assumption is not warranted for the rest of the nation.) Occupational data are particularly sensitive to definitional and procedural changes. For example, in 1890 and 1900, the federal census was taken during the month of June. In 1910 and 1930, the population was enumerated in April. In 1920, however, the census was taken in January, a slack period for farmers. The Brookings Institution estimates that more than
half a million agricultural workers were underenumerated in 1920 simply because of the change of month in scheduling the census. Furthermore, the introduction of the "labor force concept" in 1940, to replace previous definitions of "gainful employment," greatly affected occupational statistics: data for the percentage of gainfully employed older workers between 1890 and 1930 must be adjusted slightly downward to make them compatible with 1940 and subsequent census data.

Finally, the evolution of American society itself affects the comparability of census data over time. This is particularly apparent when one considers the multiplication of new occupations resulting from innovations in technology, transportation, bureaucratic practices, and large-scale industrialism. In 1890, census officials classified every occupation into one of five groups: agriculture, fishing, and mining; professional service; domestic and personal service; trade and transportation; and manufacturing and mechanics. There were thirteen major categories of occupations by 1940, including such new classifications as construction, finance, business, and government workers. The schema has been revised several times since then to reflect subsequent changes in the labor force. The growing complexity of the occupational structure becomes problematic because of its effect not only on the numerical size of any given occupational category but also on the age structure within occupational groupings.

Cataloguing deficiencies in the census statistics and citing possible methodological problems in comparing such data over time should not lead anyone to the conclusion that it is foolhardy to analyze the material available in the decennial census volumes. Rather, recognizing the data's limitations makes one wary of attaching too much significance to dramatic changes in demographic or economic rates from one decade to the next. Abrupt shifts may reflect a new direction in census definitions or procedures, not an important change in actual behavior. Therefore, what is essential is that the student of history try to grasp the meaning of long-term trends that gradually emerge over time.