Consistency of Census and Vital Registration Data on Older Americans: 1970-1990

LAURA B. SHRESTHA and SAMUEL H. PRESTON[1]

ABSTRACT

Major uncertainties about the quality of elderly population and death enumerations in the United States result from coverage and content errors in the censuses and the death registration system. This study evaluates the consistency of reported data between the two sources for the white and the African-American populations. The focus is on the older population (aged 60 and above), where mortality trends have the greatest impact on social programs and where data are most problematic. Using intercensal cohort analysis, age-specific inconsistencies between the sources are identified for two periods: 1970-1980 and 1980-1990. The U.S. data inconsistencies are examined in light of evidence in the literature regarding the nature of coverage and content errors in the data sources. Data for African-Americans are highly inconsistent in the 1970-1990 period, likely the result of age overstatement in censuses relative to death registration. Inconsistencies also exist for whites in the 1970-1980 intercensal period. We argue that the primary source of this error is an undercount in the 1970 census relative to both the 1980 census and the death registration. In contrast, the 1980-1990 data for whites, and particularly for white females, are highly consistent, far better than in most European countries.

KEY WORDS: Age misreporting; Coverage; Mortality; Census evaluation; Death registration; Data quality; Mortality crossover, United States.

1. INTRODUCTION

Conventional methods of estimating levels of mortality in more developed countries use data from two different sources. The numerators of death rates are normally counts of deaths derived from vital statistics. The denominators are usually derived from census counts of persons alive. The accuracy of calculated rates depends on the quality of data from both sources.

This paper reports the results from a test of data quality applied to United States data for two intercensal periods: 1970-1980 and 1980-1990. In particular, we examine the consistency of reported changes in the size of a cohort between two censuses and the recorded number of intercensal deaths for that cohort, with allowance for intercensal cohort migration. All data refer to the population in single years of age and separate tests are conducted for the black and white populations.

Our focus is on the older population (aged 60 and above), where mortality trends have the greatest impact on social programs (Preston 1993) and where data quality is most problematic. The white population of the United States appears to have lower death rates above age 80 than any other industrialized country (Vaupel 1993). If valid, this comparison would have important implications for evaluating the relative quality of medical systems. But the African-American population of the United States has even lower rates than the white population above age 80,

reflecting the well-known crossing over of the age patterns of mortality between the races somewhere between ages 75 and 85. Whether either set of mortality rates can be accepted at face value depends, of course, on the quality of the data. Data on blacks has elicited considerable skepticism (e.g., Zelnik 1969; Coale and Kisker 1990), although most observers appear to accept the validity of the crossover (Manton et al. 1986; McCord and Freeman 1990).

In the process of constructing new model mortality patterns for low mortality countries, Condran, Himes, and Preston (1991) report similar data quality tests for 68 intercensal periods in 18 industrialized countries. In general, consistency was very good for cohorts aged 65 at the second census (66 of 68 data sets passed the consistency check). Consistency deteriorated with age; only about half of the data sets showed consistency at age 85 and fewer than 15% did so at age 95 (Condran et al. 1991: Table 7). The United States was not among the countries included in these tests because it lacked published data on deaths by single year of age. We are now able to fill in this important gap because we have processed data tapes on each individual death registered in the United States from 1970 through 1988. (The single year death distribution for 1989 (full year) and 1990, January to March only, is estimated using published group data from the National Center for Health Statistics and the 1988 single-year death distribution. Details are provided in Appendix A.) These tapes are produced by the National Center for Health Statistics

[1] Laura B. Shrestha, The World Bank, Human Development Department, 1818 H Street, N.W., Washington, DC 20433, U.S.A.; Samuel H. Preston, Population Studies Center, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.

(NCHS) and are distributed by the Inter-University Consortium for Political and Social Research. For the years we have included, they contain approximately 50 million deaths.

2. STUDY POPULATION AND DATA

2.1 Background

Three major sources of data are utilized: (1) national-level census enumerations from the U.S. Bureau of the Census for the years 1970, 1980 and 1990; (2) annual death registration data produced by NCHS; and (3) unpublished estimates of net immigration obtained from the U.S. Bureau of the Census. While the data sources are described in more detail in Appendix A, a brief description of the data and significant adjustments is warranted.

2.2 The Census Enumerations

We utilize census tabulations which are classified by race (black/white), sex, and single years of age (open-ended at age 100). The tabulations refer to the resident population of the 50 states and the District of Columbia. Included in the enumerations are: the institutionalized population, Americans travelling abroad temporarily, and foreign citizens having their usual residence (legally or illegally) in the United States (except foreign military and diplomatic personnel). Specifically excluded are: Americans overseas for an extended period and foreign citizens temporarily visiting the U.S. The official statistics do not adjust for census undercount, e.g., the failure to find and enumerate legal residents and undocumented resident immigrants.

The term "resident population" implies that both the legal population and undocumented immigrants are included in the census tabulations. While undocumented persons were residing in the U.S. at the time of the 1970 census, it appears that only a negligible number were counted. Hence, the legal resident population approximated the total resident population in the 1970 census. In the 1980 count, however, the U.S. Bureau of the Census estimates that, for the first time ever, a significant number of undocumented persons were enumerated. Estimates indicate that the count equalled 2.06 million undocumented persons. Of this number, in the age group 60 and above, 10 thousand white males were enumerated; 19 thousand white females, 3 thousand black males, and 6 thousand black females (U.S. Bureau of the Census 1988).

The official 1970 census tabulations are known to contain errors, the most conspicuous of which is the gross overstatement of the number of persons aged 100 years or more. Although the census enumerated 106,000 persons in this age group, indirect demographic estimates indicated that the correct centenarian count should have been in the range of 3,000 to 8,000 with a preferred estimate of 4,800

(Siegel 1974; Siegel and Passel 1976; U.S. Bureau of the Census 1974). We utilize unpublished U.S. Census Bureau tabulations of the 1970 census, which correct for the centenarian overcount. Use of the corrected estimates is justified by two conditions: first, without adjustment, the excess is large enough to bias results at the oldest ages. Second, it appears that the overcount was not due to systematic misreporting of age into the centenarian population. Rather, it was the result of misunderstanding of the census form wherein individuals confused the columns intended for month of birth and year of birth (Siegel and Passel 1976).

In both the 1980 and 1990 censuses, a large number of individuals enumerated chose to write in a response to the race question as opposed to selecting one of the specified all-inclusive race categories. For the total population, 6.8 million individuals, largely of Spanish-origin, were affected in 1980, whereas the number increased to 9.3 million in the 1990 census. The official census tabulations are not directly comparable with other data sources since only the census enumerations contain a residual race category. To allow comparison with other data systems, the Census Bureau modified the 1980 and the 1990 enumerations to conform to historical categories of the racial groupings. The 1990 modification at the Census Bureau also involved "correction" for an age-related problem (for details, see Word and Spencer 1991). The decision was made to use the race-modified statistics for 1980 and 1990 from the Census Bureau for this research. The choice is justified by the sheer magnitude of individuals that would be excluded by use of the unmodified data, particularly for the white population.

2.3 Death Registration Data

The U.S. death registration data represent every death registered in the 50 states and the District of Columbia, classified by race, sex and age (single years of age to 125 +). To insure comparability with the census data, deaths of nonresidents of the United States (nonresident foreign nationals and U.S. nationals residing abroad) have been excluded.

Adjustment is made for neither under-registration of deaths nor for misreporting of characteristics on the death certificates. Two problems were identified that affected the utilization of our intended intercensal methodology. The intercensal period covers the interval from April 1 to March 31, whereas the death registration data refer to calendar years. And both the death registration and the U.S. censuses' data are reported by age at last birthday rather than by year of birth. We manage both problems by assigning deaths to triangles of time-age that correspond to "census years" beginning on April 1. For example, deaths reported in the one year interval between census date April 1, 1970 and April 1, 1971 to those aged 60 (last

birthday) at the time of the census can be classified into four categories: (1) deaths to those aged 60 in calendar year 1970; (2) deaths to those aged 60 in calendar year 1971; (3) deaths to those aged 61 in calendar year 1970; and (4) deaths to those aged 61 in calendar year 1971. Using data on the date of death from the NCHS tapes, we assigned deaths to triangles of time-age that corresponded to the census year beginning on April 1, 1970. In doing so, we assume that deaths within each triangle are evenly distributed. This assumption is necessitated by the lack of reliable birth data for most of the cohorts considered in the paper, data that could be used to apportion deaths more accurately among adjacent birth cohorts. For a more detailed description of the methodology, see Shrestha 1993.

2.4 Net Immigration Statistics

We utilize unpublished net immigration statistics obtained from the U.S. Bureau of the Census. While the quality of net immigration statistics in the U.S. is widely acknowledged to be suspect (Hill 1985), the estimation of population size at the older ages is quite robust to variations in estimates of intercensal migration. This robustness results both from the smaller flow of net migrants at the older ages relative to younger ages and from the greater magnitude of deaths as a source of decrement in the older age groups relative to changes as a result of net migration. For instance, the net immigration data list an inflow of 64 black males for the cohort aged 75 and above (in 1970) during the 1970 to 1980 decade. For comparison, over 141 thousand deaths were recorded for the same cohort.

Estimates of the flow of undocumented residents are not included in the constructed net immigration series, but will be considered in the interpretation of results. Their exclusion was precipitated by a number of factors. Estimates of the size and age-sex distributions of the illegal alien population vary widely due to insufficient data collection instruments in the U.S. But even the most exaggerated estimates of the number of undocumented migrants are minuscule relative to deaths at the older ages.

We have described a number of adjustments that we have made to the basic data: use of unpublished 1970 census tabulations because of a gross overcount of the centenarian population in the official statistics, use of race-modified tabulations of the 1980 and 1990 census, and exclusion of estimates of the undocumented alien population. In order to judge the effect of these adjustments on our results, we carried out numerous sensitivity analyses using uncorrected data. While only modest differences were observed between the results using official statistics and those with corrected tabulations (except at ages 100 and above), the intercensal cohort analyses using uncorrected data generally produced greater deviations in our final results, implying the overall appropriateness of these corrections.

3. SOURCES OF ERROR IN CENSUSES AND DEATH DATA

Errors in demographic data have been classified into coverage errors and content errors. Coverage refers to the completeness with which persons or events that fall within the defined universe of a particular data system are recorded. Content refers to the quality of information about the persons or events that are in fact recorded. Either type of error in any data source can create inconsistencies between intercensal change in cohort size and intercensal deaths. However, if both censuses and death registration suffer from the same net omission rate, then the sources will be consistent with one another; but under these circumstances, recorded death rates will also be accurate.

Identical patterns of age misreporting in censuses and death registration will not, in general, produce consistency between changes in cohort size between the censuses and recorded numbers of intercensal deaths. The reason is that, because death rates rise with age, the age distribution of deaths at older ages is older than the age distribution of population. For example, if 10% of both persons and deaths at true ages 75-79 are misreported into the age interval 80-84, then the proportionate impact on population counts will be greater than the proportionate impact on death counts. Such a pattern of age misreporting would distort death rates, and would also be visible in the consistency tests that we apply.

The Census Bureau has used demographic and statistical procedures to estimate the completeness of census coverage. Demographic procedures compare estimates of the true numbers of births minus estimated cohort deaths and migrations to census counts (see the summary in Robinson et al. 1993 and Himes and Clogg 1992). Statistical procedures match a group of individuals identified in an alternative data source (such as the Current Population Survey) to individual-level records from the Census. A third approach is to compare the Census count of older persons to the count of individuals in Medicare files.

A number of general conclusions for the old-age population were reached in the evaluation studies of the 1970 census undertaken by the U.S. Bureau of the Census (1973, 1974, 1975). First, the magnitude of net error (combination of coverage and content errors) in the old-age statistics is greater than for the younger population. Second, females exhibited higher net error rates than males, largely the result of higher levels of age misreporting. But, gross omission rates (which are only one component of net error) were higher for males. Third, levels of net error, of gross omission, and of misreporting of demographic characteristics are considerably higher for the U.S. black population than for the white. Fourth, the evidence suggests that considerable age misreporting exists in the official statistics. For example, it is interesting to note that,

for all four race-sex groups at ages 65-69 years in 1970, the estimates derived by demographic analysis suggest net census overcounts, whereas the Medicare linkage study found gross census omissions in the magnitude of 2.1% (for white females) to 12.6% (for males of the black and other races category). This comparison implies that, while these groups have gross omissions in the number of persons enumerated at ages 65-69, other larger errors (presumably, especially age overstatement among persons below age 65) are operating in the other direction to inflate the net overcount estimates at these ages. One implication is that the characteristics of a substantial part of the population reported as 65 and over in the Census relate to persons who are in fact under age 65 (U.S. Bureau of the Census 1976).

Relative to the 1970 census, the net error rates in 1980 in most of the age-race-sex groups were significantly lower. As noted by the Bureau of the Census (1988), however, results from the Post Enumeration Program (PEP) and from the 1980 Housing Unit Enumeration Duplication study affirm that a considerable proportion of the total census count, likely in excess of 1.1%, represented duplicate enumerations of individuals already in the census. Evidence implies much lower levels of duplication in earlier censuses. Thus, "regrettably, duplication receives dubious credit for part of the improvement in 1980 in net census coverage" (U.S. Bureau of the Census 1988:10).

The Census Bureau plans exhaustive evaluations of the quality of the 1990 Census, but the release of such analyses has been fragmentary to date. It does appear that the gross undercount was lower in 1980 than in 1990 (Robinson *et al.* 1993), but this may be the result of a higher degree of duplications in the 1980 census. A number of generalizations can be made regarding the pattern of net undercount in the 1990 census for the aged population. First, following its historical trend, the net error estimates for African-Americans surpass those of whites by a wide margin. The largest differential is noted for males aged 60-64. The net undercount rate for black males equals 10.3 percent, surpassing the white male estimate of 2.6 percent by 7.7 percentage points. Second, whereas undercounts are observed for all of the male aged categories, overcounts are noted in many of the female groups. Finally, as noted by Robinson *et al.* (*ibid*), the net coverage patterns are generally consistent across the last three censuses for each race-sex group.

Official death statistics produced by the National Center for Health Statistics are the basic source of annual mortality data in the United States. The figures are generally utilized without adjustment for underregistration or for misreporting of characteristics on the death certificate. It is generally assumed, however, that the death registration system is practically complete (Wilkin 1981; U.S. Bureau of the Census 1984a; National Center for Health Statistics

1968) although no national test of its comprehensiveness has been conducted since the completion of the Death Registration Area in 1933. This assumption is based on the strict legal requirements for registration as well as on the needs of survivors for proof of death in connection with burial, settling estates and collecting insurance benefits (U.S. Bureau of the Census 1984a; Wilkin 1981). Calculations by Coale and Kisker (1990), however, suggest that underregistration of deaths exists, particularly at the older ages. For the nonwhite population, for instance, registered deaths were 7% fewer than Medicare deaths for the male population aged over 80 in 1980, whereas registered female deaths were 10% fewer. These numbers, however, may be reflective of differential age reporting between the two sources, rather than of underenumeration.

The best evidence regarding the consistency of age reporting between censuses and death registration – undoubtedly the most important source of content error affecting our consistency test – matched a sample of death certificates from May to August 1960 with the 1960 census records (NCHS 1968; Hambright 1969). Although the data were collected before the time frame considered for this project, the study's findings provide insight into what may be a continuing pattern of biases present in the census and death statistics. The authors found: (1) for whites, there was fairly high agreement between the sources even with increasing age – for nonwhites, however, there was less agreement; (2) in the event of disagreement, age discrepancies for the white population between the sources were generally within one year – for nonwhites, however, the typical difference was more than one year, particularly at ages 45 and above; and (3) for whites of all ages and nonwhites aged less than 45 years, the age reported on the death certificate was typically older than that reported on the census – for nonwhites aged 45 and above, however, age reported on the death certificate was, on average, younger than on the census.

This study was unable to ascertain which data source, if either, provides the "true" age. To this end, Rosenwaike and Logue (1983) attempted to verify age reporting on the death certificate for the population aged 85 and over in the 1968 to 1972 period. The authors selected a sample of death records from those filed for decedents of extreme age in Pennsylvania and New Jersey. They then linked the individual who died to the 1900 manuscript census of population. A total of 1429 decedents were linked of whom 960 were white and 496 were non-white.

They found that age agreement of matched census records with death certificates decreased as age increased for both racial groups. Striking differences were noted between racial groups. Agreement levels for whites were high, except at ages 100 and over. For nonwhites, however, significantly lower agreement was found. The authors further note that, within race, there was little difference by sex in agreement on age.

4. AN INTERCENSAL METHODOLOGY TO EVALUATE THE QUALITY OF OLD-AGE STATISTICS

This analysis examines the extent of inconsistency in old-age U.S. data sources using an intercensal cohort methodology. The expected size of an open-ended age cohort in the second census can be estimated from its size at the first census and the intercensal deaths occurring to that cohort, after adjustment for migration (Condran, Himes and Preston 1991). Use of an open-ended category allows observation of the ratio trend while dampening error-induced extreme values at particular ages. It is insensitive to any errors of age reporting in deaths or population that occur within the population above the age that begins the open-ended age interval.

Using census enumerations and death and migration statistics for an intercensal period, intercensal cohort analysis allows us to estimate the expected size of each open-ended age cohort in the subsequent census. The previously mentioned statistics, classified by single years of age, by sex, and by two races (white, black), were utilized to calculate the following equation for the expected population at the time of the second census:

$$\hat{N}_x(2) = N_{x-10}(1) - D_{x-10}(1) + M_{x-10}(1) \qquad (1)$$

where

$\hat{N}_x(2)$ = the predicted population aged x and above at the second census, taken 10 years after the first.

$N_{x-10}(1)$ = the enumerated population aged $x - 10$ and above at time 1, the first census.

$D_{x-10}(1)$ = the intercensal deaths which had occurred to the cohort aged $x - 10$ and above (at the first census).

$M_{x-10}(1)$ = intercensal net legal immigration into the cohort aged $x - 10$ and above (at the first census).

Similarly, the expected population *at* a given age (as opposed to at age x and above) can be calculated in an analogous manner. In either circumstance, the ratio of the observed population, enumerated in the subsequent census, to the expected population, can then be calculated (after simplifying the notation and assuming net migration to be zero) as:

$$R_x = \frac{N_x(2)}{N_{x-10}(1) - D} . \qquad (2)$$

The change in the size of the cohort as measured at two successive censuses can be produced only by death or migration. A ratio of 1.00 would indicate complete consistency among the data sources. (Note that a ratio of 1.00,

while highlighting consistency, does not assure accuracy. On an individual level, for instance, if a person's age was consistently overstated by n years, the method would fail to capture the misreporting.) In fact, however, the reported count will also be affected by: (1) coverage errors in either or both censuses; (2) under- (or over-) enumeration in the death registration data and/or the immigration statistics; and (3) misreporting of characteristics (age, race, *etc.*) in any or all of the data sources (Ewbank 1981; Shryock and Siegel 1976; Condran *et al.* 1991). The ratio of observed to expected population is a useful diagnostic tool if patterns of deviation from 1.00 can be interpreted in terms of these underlying data errors. It is not a highly precise tool because different forms of error can produce the same pattern of ratios. Nevertheless, it can help discriminate among competing alternatives.

5. HOW PATTERNS OF ERROR WILL AFFECT OBSERVED/EXPECTED RATIOS

Effects of certain types of error are visible directly in the formula for the ratio itself (and have been confirmed by simulations that we have performed). To simplify the exposition, define R_x in equation (2) as the ratio of observed to expected population for age $x +$ at the second census. The following major possibilities for coverage error, and their implications for the age-pattern of ratios, can be distinguished:

1) If $N_{x-10}(1)$ and D are equally complete and $N_x(2)$ has a relative completeness level of $C(2)$, then the age pattern of ratios will be constant with age and its level will be $C(2)$.

2) If $N_x(2)$ and D are equally complete and $N_{x-10}(1)$ has a relative completeness level of $C(1)$, then the age pattern of ratios will be:
   a) Above 1.00 and rising with age if $C(1) < 1.00$
   b) Below 1.00 and falling with age if $C(1) > 1.00$.

The reason why an age trend in R_x results from this pattern of error is that a particular proportionate error in $N_{x-10}(1)$ creates increasingly larger proportionate errors in the denominator as the two offsetting terms (one positive and one negative) in the denominator grow more equal in absolute value. This equalization occurs because a higher fraction of each cohort dies during the intercensal period as age advances.

3) If $N_{x-10}(1)$ and $N_x(2)$ are equally complete and D has a relative completeness level of $C(D)$, then the age pattern of ratios will be:
   a) Above 1.00 and rising with age if $C(D) > 1.00$ (*i.e.*, if completeness of death registration exceeds the completeness of enumeration in both censuses).
   b) Below 1.00 and falling with age if $C(D) < 1.00$.

Once again, an age trend is introduced because an equal proportionate error in D will create larger proportionate errors in the denominator as its two components become more equal in absolute value.

Some of the effects of age misreporting patterns can also be understood by examining the components of this formula. Shrestha (1993) and Condran et al. (1991) introduce various errors into simulated errorless data sets typical of the current demographic conditions of the United States and the Netherlands respectively. They show that a pattern of net overstatement of age that is confined to the two censuses will produce a pattern of ratios that hovers around 1.00 until advanced ages, whereupon it falls to very low values. The reason why the ratio declines below 1.00 is, once again, that an error in one component of the denominator (in this case, inflation of $N_{x-10}(1)$ by age overstatement) introduces disproportionate effects in the denominator. Even though the rapid tapering off in the age distribution can result in $N_x(2)$ being more inflated than $N_{x-10}(1)$, eventually the inflation of the denominator exceeds that of the numerator and the ratios fall. (For an illustration, see Figure 1 of Condran et al. 1991).

Age overstatement that is confined to deaths will create a pattern of ratios that is above 1.00 and rises with age; the denominator is too low (its negative component is too large) and the proportionate deficit grows with age.

Introducing the *same* pattern of age overstatement into deaths and population figures also creates ratios that eventually rise with age. This important result is robust to the extent of error introduced (Condran et al. 1991). It reflects the fact that age distributions taper off more and more rapidly as age advances, so that the *same* percentage of persons who overstate their true age will introduce larger *percentage* errors in the reported age distributions at the very advanced ages. That is, $N_x(2)$ has a larger inflation factor than $N_{x-10}(1)$. In this case, some inflation in $N_{x-10}(1)$ is offset in its effects on the denominator by an inflation in D.

6. RESULTS

Intercensal cohort analysis was carried out for the four sex-race groups in the United States in the 1970-1980 and 1980-1990 periods. Figures 1 and 2 present the calculated ratios of the observed to expected population at selected ages by race, sex, and intercensal period.

In all race-period combinations, the age pattern of ratios is virtually the same for females and males. In all cases, the degree of inconsistency increases with age, although any systematic and significant departure from 1.00 is postponed until age 95 and beyond for whites in 1980-1990. There is clearly a discontinuity in many of these series at age 100, reflecting the idiosyncrasies of age reporting and Census Bureau adjustment procedures among centenarians.
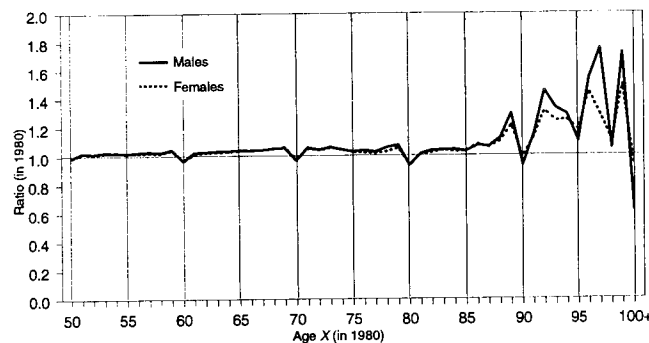

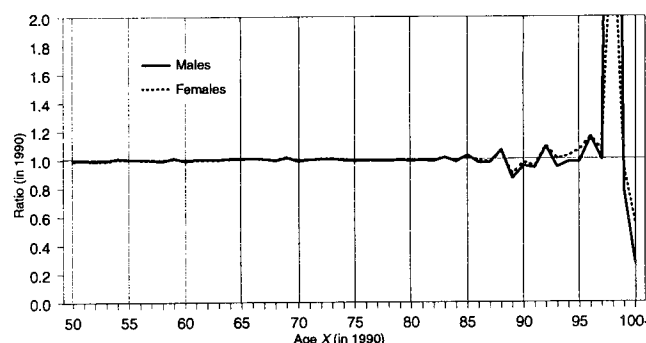**Figure 1A.** Intercensal Ratios of Observed to Expected Population: Whites, 1970-1980.

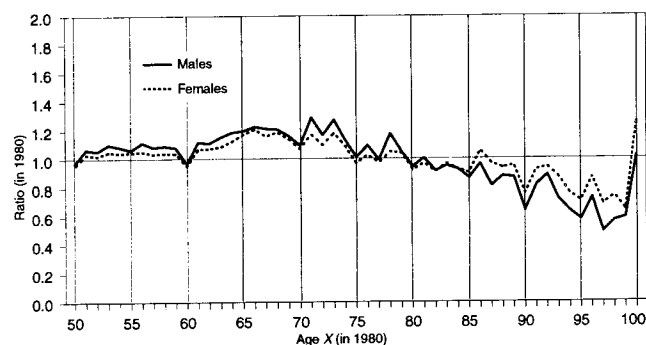**Figure 1B.** Intercensal Ratios of Observed to Expected Population: Whites, 1980-1990.

Figure 2A. Intercensal Ratios of Observed to Expected Population: Blacks, 1970-1980.
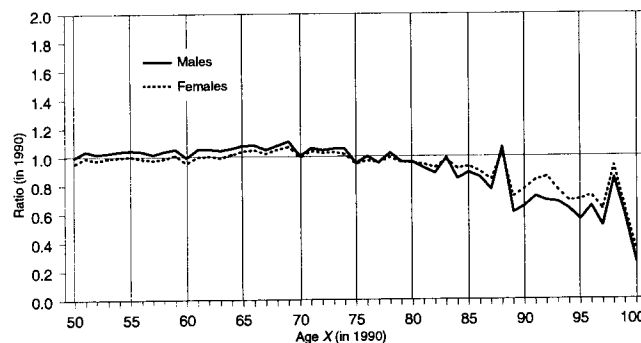

Figure 2B. Intercensal Ratios of Observed to Expected Population: Blacks, 1980-1990.

6.1 Results for Whites

6.1.1 Intercensal Period: 1970-1980

As shown in Figure 1A, the white pattern in 1970-1980 is generally above unity and rising with age (up to age 100). This pattern is consistent with several forms of data error, the two most plausible patterns of which are:

1) Undercount in the 1970 census, relative to both the 1980 census and the death registration.

2) Roughly equal probabilities of age overstatement in deaths and in both censuses.

We believe that the former explanation is more likely to be correct. If the pattern of ratios resulted from similar tendencies for age misstatement in deaths and censuses, one would expect that pattern to continue into the 1980-1990 decade, particularly since the 1980 census is involved in both comparisons. And one would not expect cultural predispositions to misstate age to disappear suddenly. But the 1980-1990 pattern of ratios for whites (Figure 1B) shows remarkable consistency, far better than that in most European countries and equivalent to the pattern of ratios found in Sweden and the Netherlands, countries with highly efficient population registers (Condran *et al.* 1991). The consistency during 1980-1990 is also much greater than that in other English-speaking countries: England and Wales, Canada, Australia and New Zealand.

A second reason for accepting the first explanation is that the Census Bureau has concluded that the 1980 census is more complete than the 1970 census (U.S. Bureau of the Census 1988; Robinson *et al.* 1993). This conclusion is partially based on demographic analysis and hence is not entirely independent of the kind of evidence that we are reviewing. However, their demographic analysis is weighted heavily towards ages that are younger than those considered here. Furthermore, the conclusion that census coverage improved is also supported by their post-enumeration program in which individuals in the census are matched against other data systems.

6.1.2 Intercensal Period: 1980-1990

As noted earlier, the 1980-1990 pattern of ratios for whites (and particularly for white females) is highly consistent, far better than in most European countries. Our investigation seemingly lends support to Vaupel's (1993) contention that the white population of the United States may have lower death rates above age 80 than any other industrialized country. But caution is in order. While our methods clearly highlight the consistency between the censuses and the death registration in 1980-1990, consistency is not equivalent to accuracy. Condran *et al.* (1991) demonstrate one situation in which a pattern of age misreporting can result in a ratio series at exactly 1.00 at all ages. Furthermore, the intercensal methodology fails to capture deliberate misreporting of age by individuals that is consistent over time. As noted by Horiuchi (1993), an initial overstatement of age – *e.g.*, to allow entrance into school or the labor force at a younger age, to avoid being drafted near the upper limit of drafting age, or to receive Social Security, Medicare, or pension payments earlier – may be followed by consistent, intentional overstatement of age. The possibility of such overstatement of age cannot be discounted although we are unable to measure it directly.

6.2 Results for Blacks

In contrast, the pattern of ratios for African-Americans is far more regular over time (see Figure 2A and 2B). The ratios begin falling around age 70 for both sexes in both periods and continue falling through higher ages (until age 100 in 1970-1980). Before age 70, ratios are typically well above unity in 1970-1980, and slightly above 1.00 for African-American males during 1980-1990.

The fact that ratios are generally higher for African-Americans at a particular age in 1970-1980 than in 1980-1990 is consistent with a relative undercount in the 1970 census. As we noted earlier, such an undercount is also likely to have occurred among whites. The undercount, however, is insufficient to explain the persistent pattern of falling ratios above age 70 in both periods. The declining ratio series for African-Americans is consistent with two principal explanations:

1) Deaths are underregistered for the African-American population relative to completeness of census coverage.

2) Age overstatement is greater in censuses than in death registration.

Coale and Kisker (1990) lean toward the former explanation. They note that populations reconstructed from deaths using variable-*r* procedures (Preston and Coale 1982) are too small relative to census counts in 1980 above age 65, suggesting relative underregistration of deaths. They also note that fewer African-American deaths are recorded at advanced ages in vital registration than in Medicare records.

However, both observations are also consistent with ages being overstated in censuses (and Medicare) relative to death registration. That such a pattern exists is strongly supported by a direct match of death certificates in 1960 to records for the same individuals in the 1960 census of population (NCHS 1968; Hambright 1969). For either males or females, the total number of deaths above age 50 when deaths are classified according to census age are within 1% of the total number of deaths when classified according to death certificate age. However, at ages 65 +, "census age" deaths are 15.4% greater than "death certificate age" deaths for females and 7.1% greater for males. At age 75 +, the disparities are 23.3% and 17.8%, respectively, and at age 85 +, 39.2% and 17.6%.

These large discrepancies in age reporting between censuses and deaths are capable of accounting for the declining pattern of ratios above age 70 that is demonstrated in Figure 2. Elo and Preston (1994) calculate the R_x values for African-Americans between 1950-1960 and 1960-1970, periods that bound the 1960 census-death certificate match. They show that, if ages at death are "corrected" to make them consistent with the age reporting in the censuses, the pattern of declining ratios is eliminated.

Reasons why African-American ages are overstated in censuses relative to deaths are not obvious. The pattern does not appear until the 1940 census, the first census after Social Security legislation was passed. At that census, a large surplus of African-American persons aged 65-69 and 70-74 appears, and a deficit of persons aged 50-64 (Elo and Preston 1994). As noted by Wolfenden (1954:56), "the disturbances were so marked in the data for Negroes that special preliminary redistributions of those populations (and deaths) between 55 and 69 were made in the preparation of the [U.S.] life tables." This surplus also appears, although in increasingly attenuated form, in more recent censuses (as shown in Figure 2). Whatever its source, we believe that the principal explanation of the large inconsistencies between censuses and death registration for the African-American population is a pattern of age overstatement in censuses relative to death registration. Such a pattern implies that recorded death rates above age 65 for African-Americans are likely to be seriously underestimated. A cross-over between black and white death rates may indeed occur at advanced ages, but basing such a conclusion on U.S. census and vital registration data is treacherous. These data are simply too inconsistent with one another to allow death rates at advanced ages to be estimated with any confidence.

7. CONCLUSION

Major uncertainties about the quality of elderly population and death enumerations in the United States result from coverage and content errors in the censuses and the death registration system. This study evaluates the consistency of reported data between the two sources for the white and the African-American populations. The focus is on the older population (aged 60 and above), where mortality trends have the greatest impact on social programs and where data are most problematic. Using intercensal cohort analysis, age-specific inconsistencies between the sources are identified for two periods, 1970-1980 and 1980-1990.

In order to evaluate what combinations of coverage completeness and age misreporting patterns would produce the empirical results, a series of simulations were carried out. The U.S. data inconsistencies are examined in light of both the simulation results and evidence in the literature

regarding the nature of coverage and content errors in the data sources.

Data for whites in the 1980-1990 intercensal period were found to be remarkably consistent. Data quality up to age 95 approaches that of Sweden and the Netherlands, countries which maintain highly efficient population registers. Less consistency was observed for whites during the 1970-1980 decade. The most likely explanation for this pattern of inconsistencies is the relative net undercount in the 1970 census combined with more complete death statistics. Consequently, mortality estimates at older ages that combine numerators from the death registration with denominators from the 1970 census are likely to overstate mortality.

A different pattern is observed in the African-American data. Above age 70, the enumerated population falls increasingly below the expected population in both 1980 and 1990. It appears that the major reason for this pattern is that ages are overstated in censuses relative to death registration. Such a pattern implies that recorded death rates at older ages for African-Americans are likely to be seriously underestimated. A mortality crossover between black and white death rates may occur at advanced ages, but basing such a conclusion on census and vital registration data is hazardous.

APPENDIX A
Source: Shrestha (1993)

Three major sources of data were utilized in this research: (1) census enumerations for 1970, 1980, and 1990; (2) official death registration data; and (3) net immigration statistics. Sources of the data and adjustments made will be described.

1.A The 1970 Census

Official tabulations of the 1970 population by basic demographic characteristics are presented in *Series B - U.S. Summary of the 1970 Census* (U.S. Bureau of the Census 1972). The official enumerations are known to

contain a number of major inaccuracies which could bias our investigation of the enumerated old-age population in the United States. The first is a conspicuous overcount of the centenarian population. Whereas 106,000 persons were enumerated in the open-ended category, indirect demographic analysis estimates the correct count to be in the range of 3,000 to 8,000 (Siegel 1974; Siegel and Passel 1976). The overcount appears to have been the result of misunderstanding of the census form rather than systematic age misreporting into the centenarian population. The second problem is the result of misclassification of the population by race in the complete-count tabulations, affecting 21,000 individuals aged 65 and above. And finally, the official count omitted over 23,000 individuals (of all ages) whose records were discovered after the initial tabulations were published.

Because of the inherent errors in the official tabulations, we utilize unpublished adjusted tabulations obtained from the U.S. Bureau of the Census. The modified statistics include corrections for the three previously mentioned problems. The data are presented by race (white, black), sex, and age (single years of age 0-94 and grouped data 95-99, 100+). To distribute the grouped data from age group 95-99 to single years of age, we used the sex-and race-specific average age distribution from the 1960 and 1980 censuses for whites, and from the 1950 and 1980 censuses for blacks (data by single years of age is not available in this age range for blacks in the 1960 census).

1.B The 1980 Census

Originally published census tabulations for 1980 were presented in *Series B - U.S. Summary of the 1980 Census* (U.S. Bureau of the Census 1983). In the 1980 census, however, a large number (about 6.8 million) of persons enumerated chose to write-in a response to the race question as opposed to selecting one of the specified all-inclusive race categories. Since only the 1980 census contained a residual race category, the official enumeration was not directly comparable with other data sources (vital registration, earlier censuses, *etc.*). The Census Bureau produced a modified file which conforms to the historical categories of the racial groupings (U.S. Bureau of the Census 1984b). The modification procedure involved macro-level reassignment of race based on detailed cross-tabulation of race and Hispanic-origin from the sample and complete-count census data. The specifics of the Census Bureau modification follow.

For the 219.8 million individuals who chose one of the 14 specified categories, no adjustment was made. Two categories of individuals, totalling 6.7 million, with write-in responses were identified: persons of Hispanic-origin (5.8 million) and persons not of Hispanic-origin (0.9 million). Separate adjustment procedures for the two groups were developed.

Those of Hispanic-origin were distributed only to the white or black categories (and not to American Indian or Asian/Pacific Islander categories). All persons of Mexican origin were reassigned as white. Persons of Puerto Rican, Cuban, and other Spanish origin were assigned to both white and black modified race groups on the basis of the distribution of the same Hispanic-origin individuals who originally specified either a white or black race on the census returns. The calculations were carried out within age-sex-county cells.

Those not of Hispanic-origin were reassigned to all three modified race groups (white, black, other) on the basis of state-specific proportions which are applied to all age-sex-county cells within the state. The proportions are based on sample data from the 1980 census. For a more detailed discussion of the modifications, see U.S. Bureau of the Census 1984b.

The modified tabulations are presented by race, sex, and single years of age (0-99; 100+). We utilize the race-modified statistics in this research, justified by the sheer magnitude of persons transferred from the residual race category to the white or black categories.

1.C The 1990 Census

Published tabulations of the 1990 Census continue to be released by the U.S. Bureau of the Census. The published statistics, however, contain a number of problems that make comparability with earlier censuses and other sources of data difficult. Three problems are apparent: racial classification of 9.3 million individuals in a residual non-specified racial category, inconsistencies in the reporting of age, and a change in allocation procedures for the 1990 census in assigning age to persons with missing data on the characteristic.

A modified 1990 census file, referred to as the MARS (Modified Age and Race Statistics) was produced at the Census Bureau to adjust for the first two problems (Word and Spencer 1991). Modification of the 1990 census was conducted at the micro-level. Hot-deck imputation procedures were utilized to assign a specific race to persons who reported themselves in the "other, not specified" racial category. The method is executed on the individual records of the 100% edited detail file from the 1990 census (Robinson, Word and Spencer 1991).

We again utilize the modified statistics, which are tabulated by race, sex, and single years of age, in this research. The decision to use the modified statistics in both 1980 and 1990 was not clear-cut. See Shrestha 1993 for a more detailed discussion.

2. The Death Registration System

National-level annual death statistics from the National Center for Health Statistics (NCHS) are utilized in this research. The data for 1970 through 1988 are extracted

from NCHS data tapes obtained from ICPSR (NCHS 1970-1988). The data are provided by race (black, white), sex, and single years of age (0-124; 125 +). Since the data tapes for calendar year 1989 and for the first three months of 1990 had yet to be released, we developed a procedure to estimate the distribution. Final mortality statistics for 1989 by race and sex were released in published form by NCHS (1992). The grouped age data was distributed to single years of age based on the 1988 death distribution within the grouped age category. Distribution to month of death was based on monthly vital statistics reports (NCHS 1989). Estimates of the death distribution in 1990 are based on monthly advance reports of mortality from NCHS (1990). The preliminary numbers were distributed to single years of age again using the 1988 distribution within the grouped age category.

As noted in the text, we adjusted the available data to correct for two problems. First, the intercensal period covers the interval from April 1 to March 31, whereas the death registration data refer to calendar years. Second, both sets of data are reported by age at last birthday rather than by year of birth. Because the census is on April 1, the latter is preferred because it identifies the birth cohort for use in cohort analysis. To adjust for these two problems, we assume that the three dimensional surface of the number of deaths in age and time is level over the interval. We do not adjust for underregistration nor for misreporting of characteristics in the death statistics.

3. Net Immigration Statistics

We utilize unpublished net immigration statistics obtained from the U.S. Bureau of the Census. The tabulations are categorized in the form of "components of change" for each of the two decades.

Age-, race-, and sex-specific net immigration was calculated on a cohort basis by use of the following equation:

Net immigration = Legal Alien Immigration + Refugees
and Parolees + Net Civilian Citizens Immigration
+ Net Puerto Rican Immigration + Net Foreign
Students Immigration + Net Movement of U.S.
Armed Forces Overseas − Legal Emigration.

Given the lack of sufficient detail in the raw data provided by the U.S. Census Bureau, a number of adjustments were required. First, the data had been provided with an early terminal age group (age 75 and above at the beginning of the decade). To distribute to five-year age groups (75-79, . . . , 95-99, 100 +), we assumed that the age-, race-, and sex-specific net immigration rate for ages 75 + remained constant in the open-ended interval beginning at age 75. This admittedly crude estimate is adequate because of small numbers of net immigrants in this age group. Second, to convert the five-year data into single years of age, we used Sprague multipliers or osculatory interpolation (Sprague 1880-81).

REFERENCES

COALE, A.J., and KISKER, E.E. (1990). Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum*, 4(1), 1-31.

CONDRAN, G.A., HIMES, C.L., and PRESTON, S.H. (1991). Old-age mortality patterns in low-mortality countries: an evaluation of population and death data at advanced ages, 1950 to present. *Population Bulletin of the United Nations*, 30, 23-60.

ELO, I.T., and PRESTON, S.H. (1994). New estimates of old-age mortality among African-Americans, 1930-1990. *Demography*, 31(3), 427-58.

EWBANK, D.C. (1981). *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis*. National Academy of Sciences, Committee on Population and Demography. Report No. 4. Washington, DC: National Academy Press.

HAMBRIGHT, T.Z. (1969). Comparison of information on death certificates and matching 1960 census records: age, marital status, nativity, and country of origin. *Demography*, 6(4), 413-24.

HILL, K. (1985). Illegal aliens: an assessment. In: Panel on Immigration Statistics. *Immigration Statistics, A Story of Neglect*. Washington, DC: National Academy Press.

HIMES, C.L., and CLOGG, C.C. (1992). An overview of demographic analysis as a method for evaluating census coverage in the United States. *Population Index*, 58(4), 587-607.

HORIUCHI, S. (1993). Personal communication dated April 20, 1993.

MANTON, K.G., STALLARD, E., and VAUPEL, J.W. (1986). Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*. 81, 635-644.

MC CORD, C., and FREEMAN, H.P. (1990). Excess mortality in Harlem. *New England Journal of Medicine*. 322, 172-177.

NATIONAL CENTER FOR HEALTH STATISTICS (1968). *Comparability of age on the death certificate and matching census record: United States - May - August 1960*; Vital and Health Statistics: Data Evaluation and Methods Research. By Thea Zelman Hambright. Series 2, No. 29.

NATIONAL CENTER FOR HEALTH STATISTICS (1970-1988). *Mortality Detail Files* (data and codebooks). Data were made available by the Inter-University Consortium for Political and Social Research, University of Michigan.

NATIONAL CENTER FOR HEALTH STATISTICS (1989). Births, marriages, divorces, and deaths for January-December 1989. *Monthly* vital statistics report. Vol. 38, Nos. 1-12. Hyattsville, Maryland: Public Health Service.

NATIONAL CENTER FOR HEALTH STATISTICS (1990). Births, marriages, divorces, and deaths for January-March 1990. *Monthly* vital statistics report. Vol. 39, Nos. 1-3. Hyattsville, Maryland: Public Health Service.

NATIONAL CENTER FOR HEALTH STATISTICS (1992). Advance report of final mortality statistics, 1989. *Monthly vital statistics report.* Vol. 40, No. 8, supp. 2. Hyattsville, Maryland: Public Health Service.

PRESTON, S.H. (1993). Demographic change in the United States, 1970-2050. *Demography and Retirement: The 21st Century,* (Sylvester Scheiber, Ed.). New York: Praeger Press.

PRESTON, S.H., and COALE, A.J. (1982). Age structure, growth, attrition, and accession: a new synthesis. *Population Index.* 48(2), 217-59.

ROBINSON, J.G., AHMED, B., DAS GUPTA, P., and WOODROW, K.A. (1993). Estimation of population coverage in the 1990 United States census based on demographic analysis. *Journal of the American Statistical Association,* 88, 1061-1079.

ROBINSON, J.G., WORD, D.L., and SPENCER, G. (1991). Uncertainty for models to translate 1990 census concepts into historical racial classifications. 1990 Decennial Census, Preliminary Research and Evaluation Memorandum (PREM) No. 81, Demographic Analysis Evaluation Project D8.

ROSENWAIKE, I., and LOGUE, B. 1983. Accuracy of death certificate ages for the extreme aged. *Demography,* 20(4), 569-585.

SHRESTHA, L.B. (1993). Age Misreporting and its Effects on Old-Age Population and Death Registration Estimates: United States, 1970-1990. Unpublished doctoral dissertation, University of Pennsylvania, Population Studies Center.

SHRYOCK, H.S., and SIEGEL, J.S. (1976). *The Methods and Material of Demography.* Orlando, Florida: Academic Press, Inc. (Harcourt Brace Jovanovich, Publishers): Studies in Population Series.

SIEGEL, J.S. (1974). Estimates of coverage of the population by sex, race, and age in the 1970 census. *Demography,* 11(1), 1-23.

SIEGEL, J.S., and PASSEL, J.S. (1976). New estimates of the number of centenarians in the United States. *Journal of the American Statistical Association.* 71, 559-566.

SPRAGUE, T.B. (1880-81). Explanation of a new formula for interpolation. *Journal of the Institute of Actuaries.* 22, 270.

UNITED STATES BUREAU OF THE CENSUS (1972). *General Population Characteristics.* 1970 Census of Population and Housing. Final Report PC(1)-B1. U.S. Summary. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1973). *The Medicare Record Check: an Evaluation of the Coverage of Persons 65 Years of Age and Over in the 1970 Census.* 1970 Census of Population: Evaluation and Research Program, PHC(E)-7. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1974). *Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis.* 1970 Census of Population: Evaluation and Research Program, PHC(E)-4. By Jacob S. Siegel. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1975). *Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match.* 1970 Census of Population: Evaluation and Research Program, PHC(E)-11. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1976). *Demographic Aspects of Aging and the Older Population in the United States.* Current Population Reports. Series P-23, No. 59. By J.S. Siegel. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1983). *General Population Characteristics.* 1980 Census of Population and Housing. Final Report PC80-1-B1. United States Summary. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1984a). *Demographic and Socioeconomic Aspects of Aging in the United States.* Current Population Reports. Series P-23, No. 138. By J.S. Siegel and M. Davidson. Washington, DC: US Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1984b). Census of Population: 1980. Race detail file. 100% count. Table IV: modified counts (OMB-consistent) by age, race, and sex. Unpublished tabulations.

UNITED STATES BUREAU OF THE CENSUS (1988). *The Coverage of Population in the 1980 Census.* 1980 Census of Population and Housing: Evaluation and Research Reports, PHC80-E4. By R.E. Fay, J.S. Passel and J.G. Robinson. Washington, DC: U.S. Government Printing Office.

VAUPEL, J.W. (1993). Verbal presentation at Research Workshop on Oldest Old Mortality, Duke University, Durham, North Carolina. March, 1993.

WILKIN, J.C. (1981). Recent trends in the mortality of the aged. *Transactions of the Society of Actuaries.* Vol. XXXIII, 11-62.

WOLFENDEN, H.H. (1954). *Population Statistics and their Compilation.* Revised Edition. The University of Chicago Press: published for the Society of Actuaries.

WORD, D.L., and SPENCER, G. (1991). Age, sex, race, and Hispanic origin information from the 1990 census: a comparison of census results with results where age and race have been modified. 1990 CHS-L-74. Draft dated August 1991.

ZELNIK, M. (1969). Age patterns of mortality of American Negroes: 1900-02 to 1959-61. *Journal of the American Statistical Association.* 64, 433-451.